

# Quantitative recovery conditions for tree-based compressed sensing

Coralia Cartis\* and Andrew Thompson\*

September 18, 2015

## Abstract

As shown in [9, 1], signals whose wavelet coefficients exhibit a rooted tree structure can be recovered – using specially-adapted compressed sensing algorithms – from just  $n = \mathcal{O}(k)$  measurements, where  $k$  is the sparsity of the signal. Motivated by these results, we introduce a simplified proportional-dimensional asymptotic framework which enables the quantitative evaluation of recovery guarantees for tree-based compressed sensing algorithms. We consider the Iterative Tree Projection (ITP) algorithm [9, 1] with a constant and a variable/practically-efficient stepsize scheme, respectively. In the context of Gaussian matrices, we apply our simplified asymptotic framework to existing worst-case analysis of ITP, which makes use of the tree-based Restricted Isometry Property (RIP). Our results have a refreshingly simple interpretation, explicitly determining a bound on the number of measurements that are required as a multiple of the sparsity. In particular, we prove that exact recovery of binary tree-based signals from noiseless Gaussian measurements is asymptotically guaranteed for ITP provided  $n \geq 115k$  (constant stepsize) and  $n \geq 683k$  (variable stepsize).

Within the same framework, we then obtain quantitative results based on a new method of analysis, recently introduced in [14], which considers the fixed points of the same ITP algorithmic variants. By exploiting the realistic average-case assumption that the measurements are statistically independent of the signal, we obtain significant quantitative improvements when compared to the tree-based RIP analysis; in this case, exact recovery of binary tree-based signals from noiseless Gaussian measurements is asymptotically guaranteed for ITP provided  $n \geq 50k$  (constant stepsize) and  $n \geq 55k$  (variable stepsize). All our results are also extended to the more realistic case in which measurements are corrupted by noise.

## 1 Introduction

Compressed sensing is motivated by the observation that many signals have an approximately sparse representation in some basis. Under this assumption, it has been proven that, to guarantee signal recovery, the sampling rate need only be proportional to the sparsity of the signal's approximation, rather than the signal dimension [16, 11]. Given an unknown signal  $x^*$  of dimension  $N$ , our aim is to recover  $x^*$  from  $n < N$  undersampled linear measurements of the form  $b = Ax^* + e$ , where  $e$  is sampling noise. Many signals have additional structure that can be exploited in the recovery process, and one such example occurs when a wavelet basis is used to represent the signal. Wavelet representations are now widely used in a variety of signal processing contexts, most notably image processing, due to the fact that piecewise smooth signals have sparse representations in wavelet bases [20]. Wavelet representations have a multi-scale tree structure, in which signals are decomposed from coarse to fine scales, with the nested support properties of wavelets inducing a parent/child relationship between wavelet coefficients at different scales. One-dimensional wavelets, for example, have a binary tree structure, in which almost all coefficients have precisely two children. Section 2.1 gives a precise characterization of the tree structures we consider here.

Since wavelets essentially work as local discontinuity detectors, signal discontinuities give rise to a chain of large coefficients along a single branch [1]. For this reason, if a particular wavelet coefficient is large, its parent wavelet coefficient is also likely to be large, which means that the large wavelet coefficients of many signals can be modelled as forming a connected subset of the whole tree which is itself a *rooted tree*. This motivates an alternative model of data simplicity: assume that the image is supported on some rooted tree of cardinality  $k$ , for some sparsity parameter  $k$ .

It was shown in [1] that it is possible to approximately perform the Euclidean projection onto the set of vectors supported on a rooted tree of given cardinality, for example by using the CSSA algorithm [2]. More recently, the present authors proposed an algorithm guaranteed to exactly calculate the projection [13]. Consequently, certain iterative projection algorithms for compressed sensing can be adapted to the tree-based

---

\*Mathematical Institute, Radcliffe Observatory Quarter, University of Oxford, Oxford, OX2 6GG, United Kingdom. E-mail: [cartis,thompson@maths.ox.ac.uk](mailto:cartis,thompson@maths.ox.ac.uk).

setting. One such algorithm proposed in [1], and also in [9], is an adaptation of the well-known Iterative Hard Thresholding (IHT) algorithm [7], which we choose to call Iterative Tree Projection (ITP). Section 2.2 gives precise details on the ITP algorithm and the two associated stepsize variants that we consider, namely, a constant stepsize scheme and a variable one, the latter being an adaptation to the tree-based setting of the practically-efficient Normalized-IHT algorithm [10, 23].

Worst-case recovery guarantees for ITP were obtained in [9, 1] by extending the ubiquitous notion of the Restricted Isometry Property [12] to the tree-based setting. Their results indicate that it suffices to take only  $n = C \cdot k$  measurements to guarantee recovery. However, the constant  $C$  has not yet been quantified: an issue of crucial importance to practitioners since it essentially determines how many measurements must be taken as a multiple of the signal sparsity. The main contribution of this paper is to obtain the first quantitative results for ITP, explicitly determining a lower bound on the constant  $C$  guaranteeing recovery. We obtain results in the context of one particular family of measurement matrices, the Gaussian ensemble, in which each entry of the matrix is i.i.d. Gaussian.

Since a Gaussian matrix is stochastic by nature, it is not possible to obtain deterministic results. However, by exploiting the remarkable concentration of measure properties of Gaussian matrices, it is possible to obtain limiting results as one lets the matrix dimensions grow. In the context of usual sparsity, Donoho introduced a *proportional-dimensional asymptotic framework* as a way of quantifying results for recovery using  $l_1$  minimization [17]. More precisely, let  $(k, n, N) \rightarrow \infty$  such that  $n/N \rightarrow \delta \in (0, 1]$  and  $k/n \rightarrow \rho \in (0, 1]$ , where  $\delta$  is the undersampling ratio and  $\rho$  is the oversampling ratio. Following this framework, limiting results were obtained in [5] for three state-of-the-art greedy algorithms including IHT, the algorithm on which ITP is based. These results, which are worst-case in nature, make use of the analysis in [8] which relies upon the RIP. More recently, by introducing a new method of analysis and by imposing an average-case assumption, the present authors obtained improved quantitative results for IHT in [14].

We now describe the main contributions of this paper.

**1) We introduce a simplified proportional growth asymptotic framework to enable quantitative comparison of recovery guarantees for tree-based compressed sensing.** The aforementioned results from [9, 1] show that tree-based compressed sensing recovery depends only upon the ratio between  $n$  and  $k$ , and is independent of  $N$ , the ambient signal dimension. This suggests that recovery results may be captured by a simplified proportional-growth asymptotic in which we dispense with the undersampling ratio  $\delta$  and consider only the oversampling ratio  $\rho$ .

**Definition 1.1 (Simplified proportional-growth asymptotic)** *We say that a sequence of problem sizes  $(k, n, N)$ , where  $0 < k \leq n \leq N$ , obeys the simplified proportional-growth asymptotic if, for some  $\rho \in (0, 1]$ ,*

$$\frac{k}{n} \rightarrow \rho \quad \text{as } (k, n, N) \rightarrow \infty.^1$$

While the common two-variable asymptotic framework leads to recovery phase transitions in the  $(\delta, \rho)$ -plane, our recovery conditions take the refreshingly simple form of a threshold  $\hat{\rho}$ , such that stable recovery is asymptotically guaranteed provided the oversampling ratio satisfies  $\rho < \hat{\rho}$ . The framework allows a direct comparison of recovery conditions for different tree-based recovery algorithms, and for different methods of analysis.

**2) We obtain the first quantitative recovery guarantees for ITP algorithms with Gaussian measurement matrices in this simplified asymptotic framework.** Our results are based upon a translation of the state-of-the-art RIP analysis in [18] to the tree-based setting, and require the derivation of upper bounds on tree-based RIP constants for Gaussian matrices in the simplified proportional-growth asymptotic. We quantify oversampling thresholds for ITP and Gaussian matrices, the precise recovery values being dependent on the ITP stepsize scheme variant used. In particular, in the case of zero noise, we prove that exact recovery of binary tree-based signals from Gaussian measurements is asymptotically guaranteed for ITP with constant stepsize provided  $n \geq 115k$  and for ITP with variable stepsize for  $n \geq 683k$ ; these results represent an improvement by at least a factor of 4 over similar asymptotic translations of the same RIP-based results for the corresponding (non-tree-based) hard-thresholding algorithms. In the case of noise, we derive *stability factors* which bound the approximation error of the output of ITP as a multiple of the noise level. The analysis in the present paper broadly follows the approach used to analyze IHT in [5, 23], but deviates from it by tightening union bound arguments by using the fact that only certain support sets (those corresponding to rooted trees) are permissible in the tree-based model.

**3) We obtain improved quantitative recovery guarantees for ITP algorithms by exploiting average-case assumptions.** We obtain results in the same framework based upon a translation of the

<sup>1</sup>Note that the only restriction that the simplified proportional-growth asymptotic places upon  $N$  is that we must have  $N \rightarrow \infty$  such that  $N \geq n$ .

*stable point* approach recently introduced by the present authors in [14] to the tree-based setting. Whereas the RIP is entirely worst-case, this alternative approach is more amenable to probabilistic analysis under the average-case (but realistic) assumption that the original signal and measurement matrix are statistically independent. Just as for the RIP analysis, the extension of the results in [14] involves the tightening of union bound arguments. Central to the analysis are large deviations results for quantities related to Gaussian matrices, which are used to bound the constituent terms of the stable point condition, employing union bounds over all permissible support sets. We obtain oversampling thresholds for the same stepsize schemes, enabling a quantitative comparison with those derived from tree-based RIP analysis. For both stepsize schemes, the incorporation of average-case assumptions leads to a significant quantitative improvement in recovery guarantees for ITP and Gaussian matrices. Namely, noiseless exact recovery of binary tree-based signals is asymptotically guaranteed for ITP provided  $n \geq 50k$  (constant stepsize) and  $n \geq 55k$  (variable stepsize); an improvement by a factor of 5 over similar asymptotic translations of stable point results for the corresponding (non-tree-based) IHT variants. We also extend our stable point recovery analysis to the case of noisy measurements, obtaining stability factors that show a substantial quantitative improvement over those derived from tree-based RIP analysis.

**Outline of the paper.** The rest of the paper is structured as follows: In Section 2, we give full technical details of the tree-based compressed sensing problem and describe in more detail the generic ITP algorithm along with two possible stepsize schemes. We describe our main results in Sections 3 and 4, first for those derived from tree-based RIP analysis (Section 3), followed by the results derived from our stable point analysis (Section 4). A discussion of all our main results then follows in Section 5. All proofs can be found in the appendix; we present the tree-based RIP analysis in Appendix A, and the stable point analysis in Appendix B. Both analyses rely crucially upon large deviations results for quantities related to Gaussian matrices (including bounds on tree-based RIP constants), and proofs of these subsidiary results can be found in Appendix C.

## 2 Problems and algorithms

### 2.1 Problem statement

Suppose we have a signal  $y^* \in \mathbb{R}^N$  which has a sparse rooted-tree representation  $x^* \in \mathbb{R}^N$  in some orthogonal wavelet basis, so that  $x^* = \Psi y^*$  where  $\Psi \in \mathbb{R}^{N \times N}$  is an orthogonal discrete wavelet transform matrix. We obtain the measurements  $b = \Phi y^* + e \in \mathbb{R}^n$ , where  $\Phi \in \mathbb{R}^{n \times N}$ , where  $e$  is the sampling noise, and where we assume  $n < N$ . Referring to  $A = \Phi \Psi^{-1} \in \mathbb{R}^{n \times N}$  from now on as the measurement matrix, we have

$$b = Ax^* + e. \quad (2.1)$$

We say that a vector  $x^*$  is *k-tree sparse* if it is supported on a rooted tree of cardinality  $k$ , and denote by  $\mathcal{T}_k$  the set of supports permitted by this model. Denoting by  $\|\cdot\|$  the Euclidean norm  $\|\cdot\|_2$ , and defining

$$\Psi(x) := \frac{1}{2} \|b - Ax\|^2, \quad (2.2)$$

we can formulate signal recovery as the following optimization problem,

$$\min_{x \in \mathbb{R}^N} \Psi(x) \quad \text{subject to} \quad \text{supp}(x) \in \mathcal{T}_k, \quad (2.3)$$

where  $\text{supp}(x)$  denotes the support of the signal  $x$ . We write  $\mathcal{P}_k$  for the (exact) Euclidean projection onto the set  $\{x : \text{supp}(x) \in \mathcal{T}_k\}$ , namely

$$\mathcal{P}_k(z) := \arg \min_{\text{supp}(x) \in \mathcal{T}_k} \|x - z\|. \quad (2.4)$$

Our analysis will consider arbitrary tree structures characterized only by the existence of a root coefficient (that is, a coefficient with no parents) and a *tree order*  $d$  defined to be the maximum number of children of any coefficient in the tree. We will at times refer to a tree of order  $d$  as a  $d$ -ary tree. The coefficients of one-dimensional wavelet transforms typically have a binary tree structure, that is tree order  $d = 2$ . The two-dimensional wavelet transforms frequently used in image processing typically form quad-trees ( $d = 4$ ). Orthogonal discrete wavelet transforms often have a particular canonical tree structure, in which every coefficient essentially has the same number of children, but this condition is never enforced in our analysis.

Our challenge, then, is to recover the wavelet representation  $x^*$  (and therefore the original signal  $y^*$ ) from the measurements (2.1), which we formally state as the following two problems.

**Problem 1 (Tree-sparse recovery from exact measurements)** Recover exactly a  $k$ -tree sparse  $x^* \in \mathbb{R}^N$  from the noiseless measurements  $b = Ax^* \in \mathbb{R}^n$ , where  $k \leq n \leq N$ .

**Problem 2 (Tree-sparse recovery from noisy measurements)** Recover a  $k$ -tree sparse  $x^* \in \mathbb{R}^N$  from the noisy measurements  $b = Ax^* + e \in \mathbb{R}^n$ , where  $k \leq n \leq N$ .

We consider the case where  $\Phi$  is chosen to be a Gaussian matrix with entries distributed i.i.d. as  $\{\Phi_{ij}\} \sim \mathcal{N}(0, 1/n)$ . The orthogonality assumption on the wavelet transform  $\Psi$  then implies that the entries of  $A$  are also distributed i.i.d. as  $\{A_{ij}\} \sim \mathcal{N}(0, 1/n)$ , i.e.  $A$  is also i.i.d. Gaussian. Assuming  $\Phi$  to be Gaussian is therefore equivalent to placing the same assumption on  $A$ , which we formalize as follows.

**Assumption 1** The measurement matrix  $A$  has i.i.d.  $\mathcal{N}(0, 1/n)$  entries.

It can be shown that  $x^*$  is the unique global solution to problem (2.3) whenever  $A$  is a Gaussian matrix [14, Sections 3 and 4.1].

**Notation.** Given some index set  $\Gamma \subseteq \{1, 2, \dots, N\}$ , we define the complement of  $\Gamma$  to be  $\Gamma^C = \{1, 2, \dots, N\} \setminus \Gamma$ . We write  $x_\Gamma$  for the restriction of the vector  $x$  to the coefficients indexed by the elements of  $\Gamma$ , and we write  $A_\Gamma$  for the restriction of the matrix  $A$  to those columns indexed by the elements of  $\Gamma$ .

## 2.2 ITP algorithms and stepsize schemes

In this section, we describe in more detail the ITP algorithm along with two possible stepsize schemes. Generically, on each iteration  $m$ , a steepest descent step, possibly with linesearch, is calculated for the objective  $\Psi$  in (2.3), namely, a move is performed from the current iterate  $x^m$  along the negative gradient of  $\Psi$ ,

$$-\nabla\Psi(x^m) = -A^T(Ax^m - b).$$

Recalling the definition of  $\mathcal{P}_k$  in (2.4), the resulting step is then projected onto the (nonconvex) constraint in (2.3) which defines the set of all vectors supported on rooted trees of cardinality  $k$ .

---

### Algorithm 2.1 Generic ITP [9, 1]

---

Inputs:  $A, b, k$ .

Initialize  $x^0 = 0$ ;  $m = 0$ .

While some termination criterion is not satisfied, do:

1.  $x^{m+1} := \mathcal{P}_k \{x^m + \alpha^m A^T(b - Ax^m)\}$ , where  $\mathcal{P}_k(\cdot)$  is defined in (2.4) and  $\alpha^m > 0$  is a stepsize.
2.  $m := m + 1$

End; output  $\hat{x} = x^m$ .

---

To avoid a situation in which the support set  $\Gamma$  is not uniquely defined, if for instance some of the coefficients are equal in magnitude, then a support set for the identical components can be selected either randomly or according to some predefined ordering. In our analysis, we will consider the possibly infinite sequence of iterates generated by ITP, though in practice a useful termination criterion such as requiring the residual to be sufficiently small, would need to be employed.

**Calculating the projection** An important property of the operator  $\mathcal{P}_k$  is that it preserves the value of selected coefficients, namely,

$$\{\mathcal{P}_k(x)\}_i := \begin{cases} x_i & i \in \Gamma \\ 0 & i \notin \Gamma \end{cases} \quad \text{where} \quad \Gamma := \arg \max_{\Gamma \in \mathcal{T}_k} \|x_\Gamma\|. \quad (2.5)$$

See [23, Lemma 6.1] for a proof of (2.5) given its definition. It follows from (2.5) that  $\mathcal{P}_k$  can be framed as an integer program with  $\{0, 1\}$  decision variables. This problem can either be solved exactly using dynamic programming [13] or approximately by solving its linear programming or Lagrangian relaxations [2, 15]. We refer the reader to [13] for further details on methods for performing the projection onto rooted trees.

**Calculating the stepsize** Two stepsize choices will be addressed in this paper: *constant stepsize*  $\alpha^m = \alpha \in (0, 1)$  for all  $m$ , which we will hereafter refer to simply as ITP [9, 1], and a *variable stepsize* scheme which we will call Normalised ITP (NITP), which adopts the same stepsize scheme as prescribed in the Normalised IHT variant of IHT algorithms proposed in [10]. The constant-stepsize ITP variant can be summarized as follows.

---

**Algorithm 2.2 ITP [9, 1]**

---

Given some  $\alpha > 0$ , on **step 1** of each iteration  $m \geq 0$  of generic ITP, set

$$\alpha^m := \alpha. \quad (2.6)$$


---

The NITP variant defined below follows [10], having the stepsize  $\alpha^m$  chosen according to an *exact linesearch* [22] when the support set of consecutive iterates stays the same, and using a *shrinkage* strategy when the support set changes, in order to ensure sufficient decrease in the objective of (2.3).

---

**Algorithm 2.3 NITP**

---

Given some  $c \in (0, 1)$  and  $\kappa > 1/(1 - c)$ , on **step 1** of each iteration  $m \geq 0$  of generic ITP, do:

**1.1. Exact linesearch.**

- (a) Set  $\Gamma^m := \text{supp}(x^m)$ .
- (b) Compute

$$\alpha^m := \frac{\|A_{\Gamma^m}^T(b - Ax^m)\|^2}{\|A_{\Gamma^m} A_{\Gamma^m}^T(b - Ax^m)\|^2}. \quad (2.7)$$

- (c) Let  $\tilde{x}^{m+1} := \mathcal{P}_k \{x^m + \alpha^m A^T(b - Ax^m)\}$ .

**1.2. Backtracking.** If  $\text{supp}(\tilde{x}^{m+1}) = \text{supp}(x^m)$ , end; output  $\alpha^m$ .

Else, while  $\alpha^m \geq (1 - c) \frac{\|\tilde{x}^{m+1} - x^m\|^2}{\|A(\tilde{x}^{m+1} - x^m)\|^2}$ , do:

- (a)  $\alpha^m := \alpha^m / (\kappa(1 - c))$ .
- (b)  $\tilde{x}^{m+1} := \mathcal{P}_k \{x^m + \alpha^m A^T(b - Ax^m)\}$ .

End; output  $\alpha^m$ .

---

In practice, the choice of  $\kappa$  in NITP constitutes a trade-off between recovery performance and computational efficiency: for optimal performance,  $\kappa$  close to 1 should be chosen, while increasing  $\kappa$  will lead to fewer shrinkage steps, making the algorithm more computationally efficient. The shrinkage strategy ensures a potentially desirable property of the NITP algorithm, namely, that provided the measurement matrix satisfies mild linear independence assumptions, it is guaranteed to converge (see Section B.1.2).

### 3 Recovery results for tree-based RIP analysis

Our first analysis relies upon a deterministic recovery condition originally given in [18] that is state-of-the-art for hard thresholding algorithms for compressed sensing. Our contribution is to extend it to the tree-based setting and then obtain from it quantitative results for Gaussian matrices. We consider an extension of the ubiquitous (asymmetric) Restricted Isometry Property (RIP) [12, 3] to the tree-based setting.

**Definition 3.1 (Tree-based RIP [9, 1])** For a given matrix  $A$ , define  $TL_s$  and  $TU_s$ , the lower and upper tree-based RIP constants of order  $s$ , to be, respectively,

$$TL_s := 1 - \min_{\emptyset \neq \text{supp}(y) \subseteq \Gamma \in \mathcal{T}_s} \frac{\|Ay\|^2}{\|y\|^2} \quad \text{and} \quad TU_s := \max_{\emptyset \neq \text{supp}(y) \subseteq \Gamma \in \mathcal{T}_s} \frac{\|Ay\|^2}{\|y\|^2} - 1. \quad (3.8)$$

#### 3.1 Tree-based RIP recovery results for ITP

Before stating a deterministic recovery result for ITP, we define two functions  $\mu^{ITP_\alpha}$  and  $\xi^{ITP_\alpha}$  which will play the role of a convergence factor and a factor controlling stability to noise.

**Definition 3.2 (Deterministic convergence and stability factor for ITP)** *Provided  $3k \leq n$ , define*

$$\mu^{ITP_\alpha} := \sqrt{3} \max\{\alpha(1 + TU_{3k}) - 1, 1 - \alpha(1 - TL_{3k})\} \quad (3.9)$$

and

$$\xi^{ITP_\alpha} := \alpha \sqrt{3(1 + TU_{2k})}, \quad (3.10)$$

where  $TU$  and  $TL$  are defined in Definition 3.1.

We have the following deterministic recovery result for ITP.

**Theorem 3.3 (Deterministic recovery result for ITP)** *Consider Problem 2. Let  $\mu^{ITP_\alpha}$  and  $\xi^{ITP_\alpha}$  be defined as in Definition 3.2. Then, provided  $\mu^{ITP_\alpha} < 1$ , the output,  $\hat{x}$ , of ITP with stepsize  $\alpha$  at iteration  $m$ , satisfies*

$$\|\hat{x} - x^*\| \leq (\mu^{ITP_\alpha})^m \|x^*\| + \frac{\xi^{ITP_\alpha}}{1 - \mu^{ITP_\alpha}} \|e\|. \quad (3.11)$$

**Proof:** See Appendix A.1. □

Though Theorem 3.3 gives a limiting bound on the approximation error, it does not necessarily imply convergence of the algorithm. In the simplified noiseless case however, the result implies convergence to  $x^*$  at a linear rate.

We derive quantitative recovery conditions for Gaussian matrices by means of upper bounds on tree-based RIP constants in the simplified proportional-growth asymptotic of Definition 1.1. We follow the broad approach used for the standard notion of RIP in [5, 6, 14], in which a union bound was performed over the maximum/minimum singular values of all  $\binom{N}{k}$  submatrices of  $A$  of size  $n \times k$ . In the present work, however, the assumed tree structure means that the number of permissible support sets for iterates of the algorithm is much diminished, which means that union bound arguments can be tightened, leading to improved quantitative results.

The number,  $|\mathcal{T}_k|$ , of permissible support sets in the  $d$ -ary tree-based framework, is bounded above by  $T(k)$ , the total number of ordered, rooted  $d$ -ary trees of cardinality  $k$ . Fortunately, a formula for  $T(k)$  is known.

**Lemma 3.4 (Tree counting result [19])** *The total number of ordered, rooted  $d$ -ary trees of cardinality  $k$  is*

$$T(k) = \frac{1}{(d-1)k+1} \binom{dk}{k}. \quad (3.12)$$

In particular, note that  $T(k)$  depends only upon the tree order  $d$  and the sparsity  $k$ , and not upon the signal length  $N$ . It is for this reason that we are able to obtain quantitative bounds in the simplified proportional-growth asymptotic, i.e. in terms of  $d$  and the variable  $\rho := \lim_{n \rightarrow \infty} \frac{k}{n}$  only.

Before defining bounds, it will be useful to define the Shannon entropy in the usual way.

**Definition 3.5 (Shannon entropy [3])** *Given  $p \in (0, 1)$ , define the Shannon entropy with base  $e$  logarithms as*

$$H(p) := -p \ln(p) - (1-p) \ln(1-p). \quad (3.13)$$

We define the following bounds on tree-based RIP constants for Gaussian matrices.

**Definition 3.6 (Tree-based RIP bounds)** *Define, for  $\rho \in (0, 1)$  and  $\lambda > 0$ ,*

$$\psi_{max}(\lambda, \rho) = \frac{1}{2} [(1 + \rho) \ln \lambda + 1 + \rho - \rho \ln \rho - \lambda] \quad (3.14)$$

and

$$\psi_{min}(\lambda, \rho) = H(\rho) + \frac{1}{2} [(1 - \rho) \ln \lambda + 1 - \rho + \rho \ln \rho - \lambda], \quad (3.15)$$

where  $H(\cdot)$  is defined in (3.13). Define  $\lambda^{max}(\rho)$  and  $\lambda^{min}(\rho)$  as the unique solution to (3.16) and (3.17) respectively:

$$\psi_{max}[\lambda^{max}(\rho), \rho] + d\rho \cdot H(d^{-1}) = 0 \quad \text{for } \lambda^{max}(\rho) > 1 + \rho; \quad (3.16)$$

$$\psi_{min}[\lambda^{min}(\rho), \rho] + d\rho \cdot H(d^{-1}) = 0 \quad \text{for } \lambda^{min}(\rho) < 1 - \rho, \quad (3.17)$$

and define  $\mathcal{TU}(\rho) = \lambda^{max}(\rho) - 1$  and  $\mathcal{TL}(\rho) = 1 - \lambda^{min}(\rho)$ .

That there exists a unique solution to (3.16) follows since  $\psi_{max}[\lambda, \rho]$  is positive for  $\lambda = 1 + \rho$ , tends to  $-\infty$  as  $\lambda \rightarrow \infty$ , and is strictly decreasing in  $\lambda$ . Similarly, that there exists a unique solution to (3.17) follows since  $\psi_{min}[\lambda, \rho]$  is positive for  $\lambda = 1 - \rho$ , tends to  $-\infty$  as  $\lambda \rightarrow \infty$ , and is strictly decreasing in  $\lambda$ . Counterparts of the bounds in Definition 3.6 for the standard notion of asymmetric RIP constants were shown to hold asymptotically for Gaussian matrices in [3]. Following their method of proof, we obtain an analogous result for tree-based RIP constants in the simplified proportional-growth asymptotic.

**Lemma 3.7 (Validity of tree-based RIP bounds)** *Suppose Assumption 1 holds and let  $\epsilon > 0$ . In the simplified proportional-growth asymptotic, we have*

$$\mathbb{P}(TU_k \geq \mathcal{TU}(\rho) + \epsilon) \rightarrow 0, \quad (3.18)$$

$$\mathbb{P}(TL_k \leq \mathcal{TL}(\rho) - \epsilon) \rightarrow 0, \quad (3.19)$$

both exponentially in  $n$ .

**Proof:** See Appendix C.

Closely following the approach in [5], we show that a naive replacement of each  $TL_{pk}$  and  $TU_{qk}$  in Definition 3.2 and Theorem 3.3 by the tree-based RIP bounds  $\mathcal{TL}(\rho)$  and  $\mathcal{TU}(\rho)$  is valid – provided the functions  $\mu_{RIP}^{ITP\alpha}$  and  $\xi_{RIP}^{ITP\alpha}$  satisfy certain properties given in Lemma A.8 of Appendix A.2. We therefore define the following asymptotic convergence and stability factors.

**Definition 3.8 (Convergence and stability factor for ITP)** *Define, for  $\rho \in (0, 1/3)$ ,*

$$\mu_{RIP}^{ITP\alpha}(\rho) := \sqrt{3} \max\{\alpha[1 + \mathcal{TU}(3\rho)] - 1, 1 - \alpha[1 - \mathcal{TL}(3\rho)]\} \quad (3.20)$$

and

$$\xi_{RIP}^{ITP\alpha}(\rho) := \alpha \sqrt{3[1 + \mathcal{TU}(2\rho)]}, \quad (3.21)$$

where  $\mathcal{TU}$  and  $\mathcal{TL}$  are given in Definition 3.6.

We finally arrive at the following asymptotic recovery result for ITP and Gaussian matrices.

**Theorem 3.9 (RIP-based recovery for ITP)** *Consider Problem 2 and suppose Assumption 1 holds. Define  $\hat{\rho}_{RIP}^{ITP\alpha}$  as the unique solution to  $\mu_{RIP}^{ITP\alpha}(\rho) = 1$ . Choose  $\epsilon \in (0, 1)$  and suppose that*

$$\rho < (1 - \epsilon)\hat{\rho}_{RIP}^{ITP\alpha}. \quad (3.22)$$

*Suppose  $\hat{x}$  is the output of ITP with stepsize  $\alpha$  at iteration  $m$ . Then*

$$\mu_{RIP}^{ITP\alpha}((1 + \epsilon)\rho) < 1, \quad (3.23)$$

and, in the simplified proportional-growth asymptotic of Definition 1.1,

$$\|\hat{x} - x^*\| \leq \left(\mu_{RIP}^{ITP\alpha}((1 + \epsilon)\rho)\right)^m \|x^*\| + \frac{\xi_{RIP}^{ITP\alpha}((1 + \epsilon)\rho)}{1 - \mu_{RIP}^{ITP\alpha}((1 + \epsilon)\rho)} \|e\|, \quad (3.24)$$

for all  $k$ -tree sparse vectors  $x^*$ , with probability tending to 1 exponentially in  $n$ .

**Proof:** See Appendix A.2. □

In the idealized case of zero measurement noise, we can deduce from Theorem 3.9 guaranteed convergence of ITP at a linear rate.

**Corollary 3.10 (RIP-based recovery for ITP: noiseless case)** *Consider Problem 1 and suppose Assumption 1 holds. Choose  $\epsilon \in (0, 1)$  and suppose that (3.22) holds, where  $\hat{\rho}_{RIP}^{ITP\alpha}$  and  $\mu_{RIP}^{ITP\alpha}(\rho)$  are defined as in Theorem 3.9. Then, in the simplified proportional-growth asymptotic, the iterates of ITP with stepsize  $\alpha$  converge to  $x^*$  at a linear rate, for all  $k$ -tree sparse vectors  $x^*$ , with probability tending to 1 exponentially in  $n$ .*

**Proof:** See Appendix A.2. □

Theorem 3.9 and Corollary 3.10 gives a continuous range of oversampling thresholds for any  $0 < \alpha < 2$ . For  $\alpha \geq 2$ , the result gives  $\hat{\rho}_{RIP}^{IHT_\alpha} = 0$  for all  $\delta \in (0, 1)$ . It is clear that  $\mu_{RIP}^{IHT_\alpha}(\rho)$  takes its minimum value when the two expressions inside the maximum in (3.20) are equal, which implies that the optimal oversampling threshold is obtained when the stepsize in the ITP algorithm is taken to be

$$\hat{\alpha} := 2/[2 + \mathcal{TU}(3\rho) - \mathcal{TL}(3\rho)]. \quad (3.25)$$

We will adopt the optimal stepsize choice  $\hat{\alpha}$  in all our numerical computations of oversampling thresholds.

### 3.2 Tree-based RIP recovery results for NITP

Turning now to the NITP stepsize scheme, we follow an analogous approach to that already described for ITP in order to finally obtain asymptotic recovery results. Omitting the deterministic results for brevity (which can be found in detail in Theorems A.5 and A.7 in Appendix A), we define the following asymptotic convergence and stability factors.

**Definition 3.11 (Convergence and stability factors for NITP)** *Define, for  $\rho \in (0, 1/3)$ ,*

$$\mu_{RIP}^{NITP_\kappa}(\rho) := \sqrt{3} \max \left\{ \frac{1 + \mathcal{TU}(3\rho)}{1 - \mathcal{TL}(\rho)} - 1, 1 - \frac{1 - \mathcal{TL}(3\rho)}{\kappa[1 + \mathcal{TU}(2\rho)]} \right\} \quad (3.26)$$

and

$$\xi_{RIP}^{NITP_\kappa}(\rho) := \frac{\sqrt{3[1 + \mathcal{TU}(2\rho)]}}{1 - \mathcal{TL}(\rho)}, \quad (3.27)$$

where  $\mathcal{TU}$  and  $\mathcal{TL}$  are defined in Definition 3.6.

We proceed to the recovery result for NITP.

**Theorem 3.12 (RIP-based recovery for NITP)** *Consider Problem 2 and suppose Assumption 1 holds. Define  $\hat{\rho}_{RIP}^{NITP_\kappa}$  as the unique solution to  $\mu_{RIP}^{NITP_\kappa}(\rho) = 1$ . Choose  $\epsilon \in (0, 1)$  and suppose that*

$$\rho < (1 - \epsilon)\hat{\rho}_{RIP}^{NITP_\kappa}. \quad (3.28)$$

Suppose  $\hat{x}$  is the output of NITP with shrinkage parameter  $\kappa$  at iteration  $m$ . Then

$$\mu_{RIP}^{NITP_\kappa}((1 + \epsilon)\rho) < 1, \quad (3.29)$$

and, in the simplified proportional-growth asymptotic,

$$\|\hat{x} - x^*\| \leq \left( \mu_{RIP}^{NITP_\kappa}((1 + \epsilon)\rho) \right)^m \|x^*\| + \frac{\xi_{RIP}^{NITP_\kappa}((1 + \epsilon)\rho)}{1 - \mu_{RIP}^{NITP_\kappa}((1 + \epsilon)\rho)} \|e\|, \quad (3.30)$$

for all  $k$ -tree sparse vectors  $x^*$ , with probability tending to 1 exponentially in  $n$ .

**Proof:** See Appendix A.2. □

In the noiseless case, we can also guarantee convergence of NITP at a linear rate.

**Corollary 3.13 (RIP-based recovery for NITP: noiseless case)** *Consider Problem 1 and suppose Assumption 1 holds. Choose  $\epsilon \in (0, 1)$  and suppose that (3.28) holds, where  $\hat{\rho}_{RIP}^{NITP_\kappa}$  and  $\mu_{RIP}^{NITP_\kappa}(\rho)$  are defined as in Theorem 3.12. Then, in the simplified proportional-growth asymptotic, the iterates of NITP with shrinkage parameter  $\kappa$  converge to  $x^*$  at a linear rate, for all  $k$ -tree sparse vectors  $x^*$ , with probability tending to 1 exponentially in  $n$ .*

**Proof:** See Appendix A.2. □



## 4 Recovery results using a tree-based stable point analysis

Our second analysis, which broadly follows the approach used to analyze IHT in [14], considers the *stable points* of ITP, a concept which can be viewed as a generalization of the notion of a fixed point to accommodate variable stepsize schemes, see [14, Section 3.1].

**Definition 4.1 (Stable points of generic ITP)** *Given  $\underline{\alpha} > 0$  and an index set  $\Gamma \in \mathcal{T}_k$ , we say  $\bar{x} \in \mathbb{R}^N$  is an  $\underline{\alpha}$ -stable point of generic ITP on  $\Gamma$  if  $\text{supp}(\bar{x}) \subseteq \Gamma$  and*

$$\{A^T(b - A\bar{x})\}_\Gamma = 0 \quad \text{and} \quad (4.31)$$

$$\|\bar{x}_{\Gamma \setminus \Omega}\| \geq \underline{\alpha} \|A_{\Omega \setminus \Gamma}^T(b - A\bar{x})\| \quad \forall \Omega \in \mathcal{T}_k. \quad (4.32)$$

We will require the following assumption for the deterministic result given next.

**Assumption 2** *The columns of  $A$  are in  $2k$ -general position, namely any collection of  $2k$  of its columns are linearly independent.*

Assumption 2 is a typical (weak) assumption in compressed sensing, and which guarantees a unique solution to Problem 1. We denote by  $A_\Gamma^\dagger$  the Moore-Penrose pseudoinverse  $(A_\Gamma^T A_\Gamma)^{-1} A_\Gamma^T$ , which is well-defined under Assumption 2.

We next state the *stable point condition*, that is, a necessary condition for the existence of a stable point on a given support. It will help to define  $\Lambda \in \mathcal{T}_k$  to be the support of the original signal, namely

$$\Lambda := \text{supp}(x^*). \quad (4.33)$$

**Theorem 4.2 (Stable point condition)** *Consider Problem 2. Suppose Assumption 2 holds and suppose there exists an  $\underline{\alpha}$ -stable point of generic ITP on some  $\Gamma$  such that  $\Gamma \neq \Lambda$ . Then*

$$\left\| A_\Gamma^\dagger A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^* \right\| + \left\| A_\Gamma^\dagger e \right\| \geq \underline{\alpha} \left\{ \left\| A_{\Lambda \setminus \Gamma}^T (I - A_\Gamma A_\Gamma^\dagger) A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^* \right\| - \left\| A_{\Lambda \setminus \Gamma}^T (I - A_\Gamma A_\Gamma^\dagger) e \right\| \right\}, \quad (4.34)$$

where  $\Lambda$  is defined in (4.33).

**Proof:** See Appendix B.1.1. □

Note that Theorem 4.2 applies to both ITP (with  $\underline{\alpha} = \alpha$ ) and NITP (with  $\underline{\alpha}$  being a lower bound on the stepsize  $\alpha^m$ ).

### 4.1 Tree-based stable-point recovery results for ITP

We first consider ITP with constant stepsize  $\alpha$  and give a deterministic condition guaranteeing convergence of ITP to some  $\alpha$ -stable point in terms of the tree-based RIP.

**Theorem 4.3 (ITP convergence)** *Consider Problem 2. Suppose that Assumption 2 holds, and suppose that the stepsize in ITP satisfies*

$$\alpha < \frac{1}{1 + TU_{2k}}, \quad (4.35)$$

where  $TU$  is defined in (3.1). Then ITP with stepsize  $\alpha$  converges to an  $\alpha$ -stable point  $\bar{x}$  of ITP.

**Proof:** See Appendix B.1.2. □

Theorem 4.3 can be quantified probabilistically for Gaussian measurement matrices using Lemma 3.7 (see Lemma B.6 in Appendix B.2.1). To obtain recovery results for ITP, namely, to ensure convergence of ITP to the original stable point/signal, we are left with analysing the stable point condition in Theorem 4.2 to guarantee that all stable points (which are candidate limit points of the ITP iterates) are ‘close’ to the original signal. While it would be possible to analyse the stable point condition using the tree-based RIP, we take a different approach. The stable point condition is especially amenable to probabilistic analysis for Gaussian matrices under the average-case (but realistic) assumption that the original signal and measurement matrix are statistically independent.

**Assumption 3** *The original signal  $x^*$  and the measurement matrix  $A$  are statistically independent.*

The crucial independence assumption will allow us to obtain better quantitative results than could be achieved through the purely worst-case RIP-based analysis of Section 3. However, it is worth noting that independence is the only average-case assumption we invoke: we assume nothing further about the coefficient values of  $x^*$ . In keeping with the spirit of average-case analysis, we also assume that the noise is Gaussian and independent of both  $A$  and  $x^*$ , which we formalize as follows.

**Assumption 4** *The noise vector  $e$  has i.i.d. Gaussian entries  $e_i \sim N(0, \sigma^2/n)$ , independently of  $A$  and  $x^*$ .*

Note that, under Assumption 4,  $\mathbb{E}\|e\|^2 = \sigma^2$ , so that  $\|e\| \approx \sigma$ .

Assumption 2 is satisfied with probability 1 by a Gaussian matrix, see [14, Section 4.1], and so may now be replaced with Assumption 1.

Under Assumptions 1, 3 and 4, each of the terms in (4.34), viewed as a Rayleigh quotient over  $\|x_{\Lambda \setminus \Gamma}\|^2$ , is distributed according to either the  $\chi^2$  or the  $\mathcal{F}$  distribution. We write  $\chi_s^2$  for the (univariate)  $\chi^2$ -distribution with  $s \geq 1$  degrees of freedom. Furthermore, if  $P \sim \frac{1}{s}\chi_s^2$  and  $Q \sim \frac{1}{t}\chi_t^2$  are independent random variables, we say that  $P/Q$  follows the  $\mathcal{F}$ -distribution, and we write  $P/Q \sim \mathcal{F}(s, t)$ . The following lemma, which was proved in [14], gives the precise distributions.

**Lemma 4.4 (Distribution results for the stable point condition [14, Lemma 4.4])** *Suppose Assumptions 1, 3 and 4 hold, and let  $\Gamma$  be an index set of cardinality  $k$ , where  $k < n$ . Then*

$$\frac{\|A_\Gamma^\dagger A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}\|^2}{\|x_{\Lambda \setminus \Gamma}\|^2} = F_\Gamma, \quad \text{where } F_\Gamma \sim \frac{k}{n-k+1} \mathcal{F}(k, n-k+1); \quad (4.36)$$

$$\frac{\|A_{\Lambda \setminus \Gamma}^T (I - A_\Gamma A_\Gamma^\dagger) A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}\|^2}{\|x_{\Lambda \setminus \Gamma}\|^2} \geq \left(\frac{n-k}{n}\right)^2 \cdot R_\Gamma^2, \quad \text{where } R_\Gamma \sim \frac{1}{n-k} \chi_{n-k}^2; \quad (4.37)$$

$$\|A_\Gamma^\dagger e\| \leq \sigma \cdot \sqrt{G_\Gamma}, \quad \text{where } G_\Gamma \sim \frac{k}{n-k+1} \mathcal{F}(k, n-k+1); \quad (4.38)$$

$$\|A_{\Lambda \setminus \Gamma}^T (I - A_\Gamma A_\Gamma^\dagger) e\| \leq \sigma \sqrt{\frac{k(n-k)}{n^2} \cdot (S_\Gamma)(T_\Gamma)}, \quad \text{where } S_\Gamma \sim \frac{1}{n-k} \chi_{n-k}^2, \quad T_\Gamma \sim \frac{1}{k} \chi_k^2. \quad (4.39)$$

Recalling the stable point condition, we wish to show that all stable points are ‘close’ to the original signal, which can be achieved by bounding each of the constituent terms over all permissible support sets. We can make an analogy with the tree-based RIP, where upper bounds on tree-based RIP constants are obtained in the simplified proportional-growth asymptotic by union bounding the tail probabilities of extreme singular values of submatrices of  $A$  corresponding to permissible support sets. Similarly, large deviation bounds over  $|\mathcal{T}_k|$  instances of  $\chi^2$  and  $\mathcal{F}$  distributed random variables can be derived in the same asymptotic framework. One can view the resulting bounds as a kind of ‘independent RIP’ [14], where the assumption of independence between the measurement matrix and the original signal allows the tightening of bounds on Rayleigh quotients. Such an analysis is only possible if matrix-vector independence can be assumed, which is the case for the stable point condition (4.34). We define three tail bound functions.

**Definition 4.5 ( $\chi^2$  tail bounds)** *Let  $\rho \in (0, 1)$  and  $\lambda \in (0, 1]$ . Let  $\mathcal{TU}(\rho, \lambda)$  be the unique solution to*

$$\nu - \ln(1 + \nu) = \frac{2d\rho \cdot H(d^{-1})}{\lambda} \quad \text{for } \nu > 0, \quad (4.40)$$

*and let  $\mathcal{TIL}(\rho, \lambda)$  be the unique solution to*

$$-\nu - \ln(1 - \nu) = \frac{2d\rho \cdot H(d^{-1})}{\lambda} \quad \text{for } \nu \in (0, 1), \quad (4.41)$$

*where  $H(\cdot)$  is defined in (3.13).*

That  $\mathcal{TU}$  is well-defined follows since the left-hand side of (4.40) is zero at  $\nu = 0$ , tends to infinity as  $\nu \rightarrow \infty$ , and is strictly increasing on  $\nu > 0$ . Similarly,  $\mathcal{TIL}$  is well-defined since the left-hand side of (4.41) is zero at  $\nu = 0$ , tends to infinity as  $\nu \rightarrow 1$ , and is strictly increasing on  $\nu \in (0, 1)$ .

**Definition 4.6 ( $\mathcal{F}$  tail bound)** *Let  $\rho \in (0, 1/2]$ . Let  $\mathcal{TIF}(\rho)$  be the unique solution in  $f$  to*

$$\ln(1 + f) - \rho \ln f = 2d\rho \cdot H(d^{-1}) + H(\rho) \quad \text{for } f > \frac{\rho}{1-\rho}, \quad (4.42)$$

*where  $H(\cdot)$  is defined in (3.13).*

That  $\mathcal{TLF}$  is well-defined follows since the left-hand side of (4.42) is equal to  $H(\rho)$  at  $f = \rho/(1 - \rho)$ , tends to infinity as  $f \rightarrow \infty$ , and is strictly increasing on  $f > \rho/(1 - \rho)$ .

**Lemma 4.7 (Tree-based large deviations result for  $\chi^2$ )** *Let  $l \in \{1, \dots, n\}$  and let the random variables  $X_l^i \sim \frac{1}{l}\chi_l^2$  for all  $i \in S_n$ , where  $|S_n| = T(k)$ , and let  $\epsilon > 0$ . In the simplified proportional growth asymptotic, let  $l/n \rightarrow \lambda \in (0, 1]$ . Then*

$$\mathbb{P} \left\{ \bigcup_{i \in S_n} [X_l^i \geq 1 + \mathcal{TU}(\rho, \lambda) + \epsilon] \right\} \rightarrow 0 \quad (4.43)$$

and

$$\mathbb{P} \left\{ \bigcup_{i \in S_n} [X_l^i \leq 1 - \mathcal{TIL}(\rho, \lambda) - \epsilon] \right\} \rightarrow 0, \quad (4.44)$$

exponentially in  $n$ , where  $\mathcal{TU}(\rho, \lambda)$  and  $\mathcal{TIL}(\rho, \lambda)$  are defined in (4.40) and (4.41) respectively.

**Proof:** See Appendix C. □

**Lemma 4.8 (Tree-based large deviations results for  $F$ )** *Let the random variables  $X_n^i \sim \frac{k}{n-k+1} \mathcal{F}(k, n-k+1)$  for all  $i \in S_n$ , where  $|S_n| = T(k)$ , and let  $\epsilon > 0$ . In the simplified proportional growth asymptotic,*

$$\mathbb{P} \left\{ \bigcup_{i \in S_n} [X_n^i \geq \mathcal{TLF}(\rho) + \epsilon] \right\} \rightarrow 0, \quad (4.45)$$

exponentially in  $n$ , where  $\mathcal{TLF}(\rho)$  is defined in (4.42).

**Proof:** See Appendix C. □

We define oversampling thresholds for ITP algorithms in terms of the above tail bounds.

**Definition 4.9 (Stable point recovery oversampling threshold for ITP)** *Define  $\hat{\rho}_{SP}^{ITP}$  to be the unique solution to*

$$\frac{\sqrt{\mathcal{TLF}(\rho)}}{(1 - \rho)[1 - \mathcal{TIL}(\rho, 1 - \rho)]} = \frac{1}{1 + \mathcal{TU}(2\rho)} \quad \text{for } \rho \in (0, 1/2], \quad (4.46)$$

where  $\mathcal{TLF}$  is defined in (4.42),  $\mathcal{TIL}$  is defined in (4.41) and  $\mathcal{TU}$  is defined in Definition 3.6.

The oversampling threshold (4.46) is a counterpart of the phase transitions given in [14, Section V.1] for IHT algorithms, with the only changes being the switch to tree-based tail bounds and the disappearance of the  $\delta$  variable. A proof that (4.46) admits a unique solution proceeds analogously to the one given for the counterpart phase transitions in [14, Section V.1]. Next, we define a function  $\xi_{SP}^{ITP\alpha}(\rho)$  which will represent a stability factor in our results, bounding the approximation error of the output of ITP as a multiple of the noise level  $\sigma$ .

**Definition 4.10 (Stability factor for ITP)** *Consider Problem 2. Given  $\rho \in (0, 1/2]$  and  $\underline{\alpha} > 0$ , provided*

$$\rho < \hat{\rho}_{SP}^{ITP}, \quad (4.47)$$

define

$$a(\rho) := \frac{\sqrt{\mathcal{TLF}(\rho)} + \alpha \sqrt{\rho(1 - \rho)[1 + \mathcal{TU}(\rho, 1 - \rho)][1 + \mathcal{TU}(\rho, \rho)]}}{\alpha(1 - \rho)[1 - \mathcal{TIL}(\rho, 1 - \rho)] - \sqrt{\mathcal{TLF}(\rho)}}, \quad (4.48)$$

and

$$\xi_{SP}^{ITP\alpha}(\rho) := \sqrt{\mathcal{TLF}(\rho)[1 + a(\rho)]^2 + [a(\rho)]^2}, \quad (4.49)$$

where  $\mathcal{TLF}$  is defined in (4.42), and where  $\mathcal{TU}$  and  $\mathcal{TIL}$  are defined in (4.40) and (4.41) respectively.

Note that (4.47) ensures that the denominator in (4.48) is strictly positive and that  $a(\rho)$  is therefore well-defined. We proceed to our recovery result for constant stepsize ITP.

We are now ready to merge the quantified convergence and stable point conditions into a recovery result for ITP.

**Theorem 4.11 (Stable point recovery for ITP)** Consider Problem 2 and suppose Assumptions 1, 3 and 4 hold. If (4.47) holds and the stepsize  $\alpha$  satisfies

$$\frac{\sqrt{\mathcal{TIF}(\rho)}}{(1-\rho)[1-\mathcal{TIL}(\rho, 1-\rho)]} < \alpha < \frac{1}{1+\mathcal{TU}(2\rho)}, \quad (4.50)$$

then, in the simplified proportional-growth asymptotic of Definition 1.1, ITP with stepsize  $\alpha$  converges to  $\bar{x}$  such that

$$\|\bar{x} - x^*\| \leq \xi_{SP}^{ITP\alpha}(\rho) \cdot \sigma, \quad (4.51)$$

with probability tending to 1 exponentially in  $n$ .

**Proof:** See Appendix B.2.1. □

In the special case of Problem 1, the same oversampling threshold guarantees exact recovery of the underlying signal  $x^*$ .

**Corollary 4.12 (Stable point recovery for ITP: noiseless case)** Consider Problem 1. Suppose Assumptions 1 and 3 hold, suppose that (4.47) holds, and suppose that  $\alpha$  satisfies (4.50). Then, in the simplified proportional-growth asymptotic, ITP with stepsize  $\alpha$  converges to  $x^*$  with probability tending to 1 exponentially in  $n$ .

**Proof:** See Appendix B.2.1. □

## 4.2 Tree-based stable-point recovery results for NITP

We now turn our attention to NITP, and for brevity, we present straightaway the recovery results (that use the oversampling threshold and stability factor given next). We note that the recovery results below are obtained similarly as for ITP: by analysing the stable point condition in Theorem 4.2 using the probabilistic results in Lemma 4.4, and providing a deterministic convergence condition (Theorem B.4) that we can quantify in the simplified proportional framework; we refer the reader to Appendix B for details.

**Definition 4.13 (Stable point recovery oversampling threshold for ITP)** Define  $\hat{\rho}_{SP}^{NITP\kappa}$  to be the unique solution to

$$\frac{\sqrt{\mathcal{TIF}(\rho)}}{(1-\rho)[1-\mathcal{TIL}(\rho, 1-\rho)]} = \frac{1}{\kappa[1+\mathcal{TU}(2\rho)]} \quad \text{for } \rho \in (0, 1/2], \quad (4.52)$$

where  $\mathcal{TIF}$  is defined in (4.42),  $\mathcal{TIL}$  is defined in (4.41) and  $\mathcal{TU}$  is defined in Definition 3.6.

A proof that (4.52) admits a unique solution proceeds analogously to the one given for the counterpart phase transitions in [14, Section V.1]. We define the following stability factor for NITP.

**Definition 4.14 (Stability factor for NITP)** Consider Problem 1. Given  $\rho \in (0, 1/2]$ , provided

$$\rho < \hat{\rho}_{SP}^{NITP\kappa}, \quad (4.53)$$

define

$$a(\rho) := \frac{\sqrt{\mathcal{TIF}(\rho)} + \{\kappa[1+\mathcal{TU}(2\rho)]\}^{-1}\sqrt{\rho(1-\rho)[1+\mathcal{TIU}(\rho, 1-\rho)][1+\mathcal{TIU}(\rho, \rho)]}}{(1-\rho)\{\kappa[1+\mathcal{TU}(2\rho)]\}^{-1}[1-\mathcal{TIL}(\rho, 1-\rho)] - \sqrt{\mathcal{TIF}(\rho)}}, \quad (4.54)$$

and

$$\xi_{SP}^{NITP\kappa}(\rho) := \sqrt{\mathcal{TIF}(\rho)[1+a(\rho)]^2 + [a(\rho)]^2}, \quad (4.55)$$

where  $\mathcal{TIF}$  is defined in (4.42), where  $\mathcal{TIU}$  and  $\mathcal{TIL}$  are defined in (4.40) and (4.41) respectively, and where  $\mathcal{TU}$  is defined in Definition 3.6.

**Theorem 4.15 (Stable point recovery for NITP)** Consider Problem 2, suppose Assumptions 1, 3 and 4 hold, and suppose (4.53) holds. Then, in the simplified proportional-growth asymptotic, NITP with shrinkage parameter  $\kappa$  converges to  $\bar{x}$  such that

$$\|\bar{x} - x^*\| \leq \xi_{SP}^{NITP\kappa}(\rho) \cdot \sigma, \quad (4.56)$$

with probability tending to 1 exponentially in  $n$ .

**Proof:** See Appendix B.2.2. □

In the case of Problem 1, Theorem 4.15 also simplifies to an exact recovery result.

**Corollary 4.16 (Stable point recovery for NITP: noiseless case)** *Consider Problem 1. Suppose Assumptions 1 and 3 hold and suppose that (4.53) holds. Then, in the simplified proportional-growth asymptotic, NITP with shrinkage parameter  $\kappa$  converges to  $x^*$  with probability tending to 1 exponentially in  $n$ .*

**Proof:** See Appendix B.2.2. □

## 5 Discussion of recovery results

### 5.1 Tree-based RIP recovery results

**Noiseless case.** The oversampling thresholds for ITP and NITP defined in Corollaries 3.10 and 3.13 are displayed in Figure 1(a) for different tree orders  $d$ . For binary trees, for example, we have  $\hat{\rho}_{RIP}^{ITP} \approx 0.00875$  for ITP and  $\hat{\rho}_{RIP}^{NITP} \approx 0.00146$  for NITP (taking  $\kappa = 1.1$  for the shrinkage parameter in NITP). In both cases, exact recovery in the noiseless case is asymptotically guaranteed provided the limiting value of the ratio  $\rho$  is less than the given threshold. We see a measured deterioration in the results for higher tree orders: the corresponding thresholds for quad-trees ( $d = 4$ ) – which arise in image analysis using 2D wavelets – are 0.00705 and 0.00123 for ITP and NITP respectively. Figure 1(b) shows the inverse of the oversampling ratio, which indicates the number of measurements required by the analysis as a multiple of the sparsity. We find, for binary trees, that  $n \geq 115k$  measurements guarantees recovery by ITP, while  $n \geq 683k$  measurements guarantees recovery by NITP. Provided the oversampling thresholds are respected, convergence to the original signal is guaranteed at a linear rate. The quantities  $\mu_{RIP}^{ITP}(\rho)$  and  $\mu_{RIP}^{NITP}(\rho)$  represent guaranteed bounds on the convergence rate for each variant.

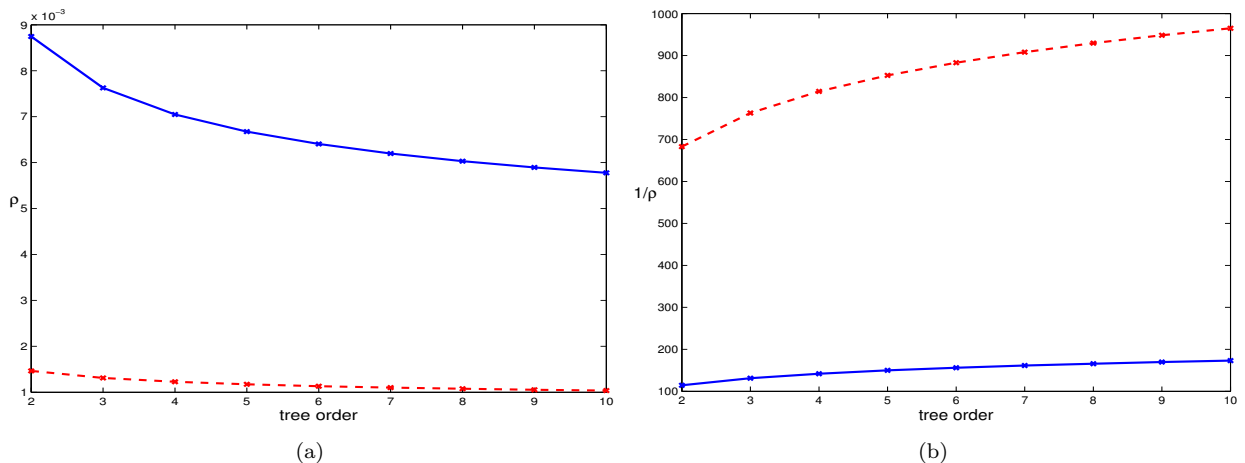


Figure 1: (a) Critical  $\rho$ -values for different tree orders from tree-based RIP analysis: ITP – unbroken; NITP – dashed. (b) Corresponding oversampling factors (reciprocals of  $\hat{\rho}$ ).

While in the present paper we have dispensed with the undersampling ratio  $\delta = n/N$ , we may also frame our results in the  $(\delta, \rho)$  asymptotic in order to make a comparison with analogous results derived in the non-tree-based setting for IHT based upon the standard notion of RIP [23]. Since there is no dependence upon  $\delta$  in our case, the phase transitions we obtain are simply horizontal lines in the  $(\delta, \rho)$ -plane. Exact recovery phase transitions for binary trees are displayed in Figure 2 alongside the phase transitions derived in [14]: recovery is guaranteed asymptotically beneath the respective curves. We observe that the switch to the tree-based setting leads to significantly improved results, especially for small  $\delta$ .

**Extension to noise.** In the case where measurements are contaminated by noise, exact recovery of the original signal is an unrealistic aim. However, provided the limiting value of the ratio  $\rho$  falls below the respective oversampling threshold, Theorems 3.9 and 3.12 give bounds on the limiting approximation error. More precisely, the results state that the limiting approximation error of the iterates of ITP/NITP is asymptotically bounded by some known stability factor multiplied by the noise level  $\sigma$ . However, neither result necessarily implies convergence of the algorithm in the case of noise. The Figure 3 plots the noise stability factor  $\xi(\rho)/[1 - \mu(\rho)]$  for binary trees, for each of the two stepsize schemes considered ( $\kappa = 1.1$  for

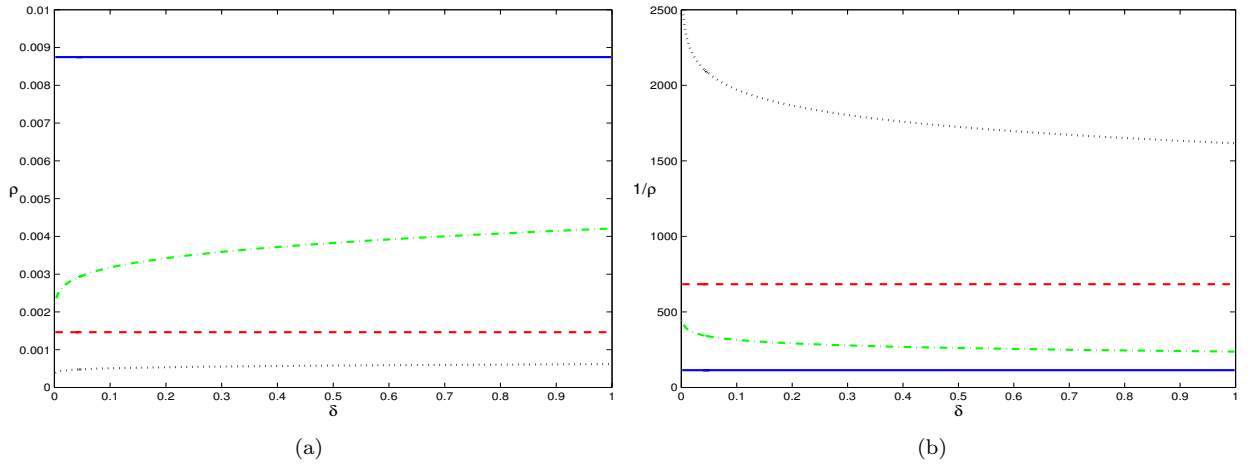


Figure 2: (a) Phase transitions from RIP analysis in the  $(\delta, \rho)$  framework for binary trees (ITP – unbroken; NITP – dashed) and non-tree-based (IHT – dash-dot; NIHT – dotted). (b) Corresponding inverses of the phase transition.

NITP). In keeping with [3], [5] and [14], we observe that the stability factor tends to infinity as the transition point is reached, *i.e.*  $\xi(\rho)/[1 - \mu(\rho)] \rightarrow \infty$  as  $\rho \rightarrow \hat{\rho}$ . For both ITP and NITP, given any value of  $\rho$  for which the stability factors derived in this paper are defined, they are always lower than the corresponding stability factors derived from analysis of IHT based upon the standard RIP [18]; see [23, Section 2.4] for a comparison.

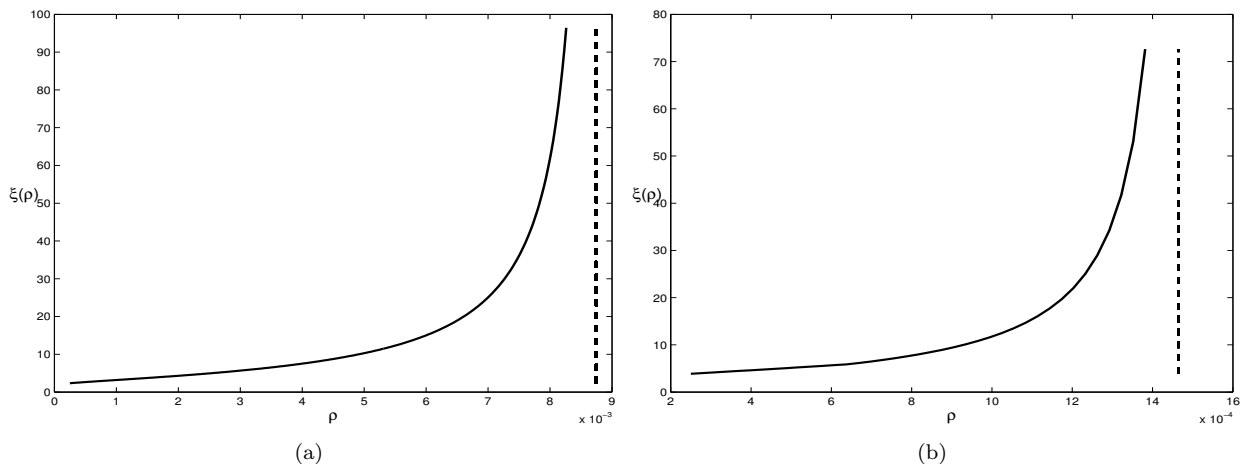


Figure 3: Plot of the stability factor  $\xi(\rho)/[1 - \mu(\rho)]$  from tree-based RIP analysis for binary trees: (a) ITP; (b) NITP.

## 5.2 Recovery results from the tree-based stable point analysis

**Noiseless case.** The oversampling thresholds for ITP and NITP defined in Corollaries 4.12 and 4.16 are displayed in Figure 4(a) for different tree orders  $d$ . For binary trees, we have  $\hat{\rho}_{RIP}^{ITP} \approx 0.0202$  for ITP and  $\hat{\rho}_{RIP}^{NITP} \approx 0.0184$  for NITP, and the corresponding thresholds for quad-trees ( $d = 4$ ) are 0.0147 and 0.0134 respectively. Figure 4(b) shows the inverse of the oversampling ratio: we find, for binary trees, that  $n \geq 50k$  measurements guarantees recovery by ITP, while  $n \geq 55k$  measurements guarantees recovery by NITP. The same exact recovery thresholds for binary trees are presented in the form of phase transitions in the  $(\delta, \rho)$  asymptotic in Figure 5, alongside the phase transitions for IHT/NIHT derived in [14]. Again, we observe improved results by switching to the tree-based setting, especially for small  $\delta$ .

Comparing the oversampling thresholds derived from the stable point analysis (Figure 4) with those derived from tree-based RIP analysis (Figure 1), we observe a significant quantitative improvement for both algorithm variants, by over a factor of 10 for NITP in fact for all tree orders under consideration. We have

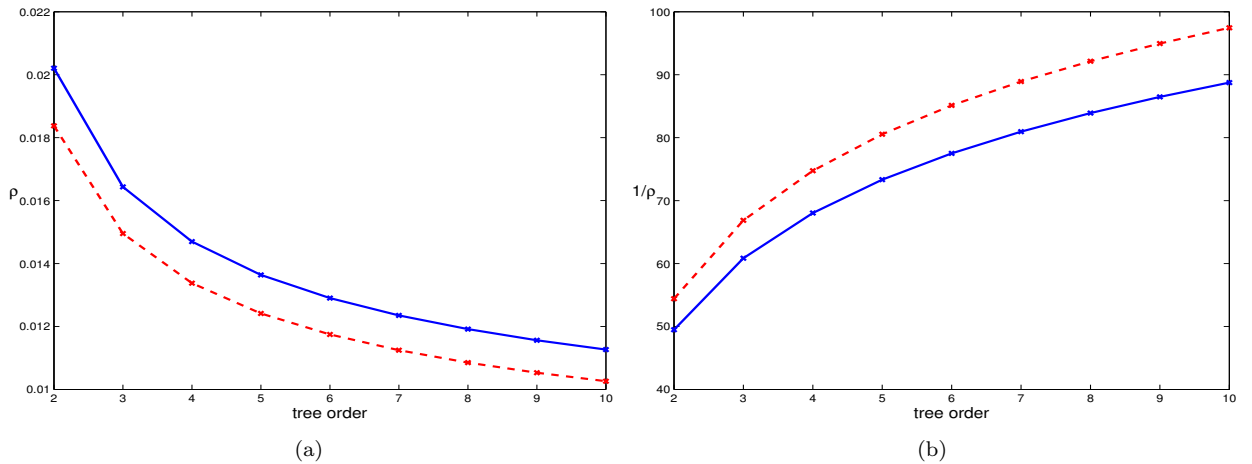


Figure 4: (a) Critical  $\rho$ -values for different tree orders from stable point analysis: ITP – unbroken; NITP – dashed. (b) Corresponding oversampling factors (reciprocals of  $\hat{\rho}$ ).

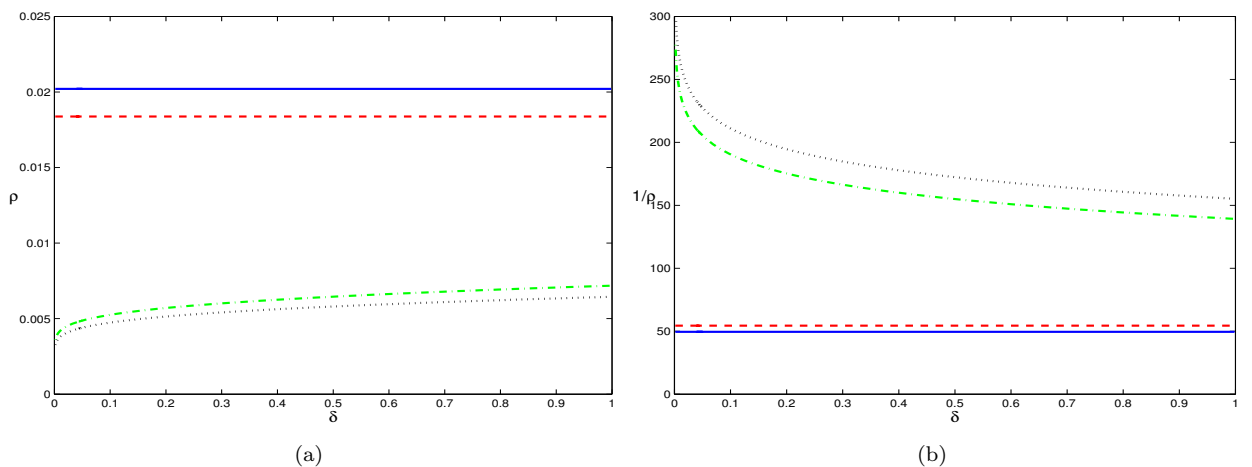


Figure 5: (a) Phase transitions from stable point analysis in the  $(\delta, \rho)$  framework for binary trees (ITP – unbroken; NITP – dashed) and non-tree-based (IHT – dash-dot; NIHT – dotted). (b) Corresponding inverses of the phase transition.

obtained improved oversampling thresholds by exploiting average-case assumptions, and we should point out the difference between the results in Sections 5.1 and 5.2. The tree-based RIP results are worst-case in nature: given a sequence of randomly generated Gaussian matrices, it is asymptotically guaranteed that ITP/NITP will in fact recover an accurate approximation to *any*  $k$ -tree sparse signal vector. On the other hand, the results derived from our stable point analysis have a more average-case flavour: given a sequence of randomly generated Gaussian measurement matrices *along with* a sequence of signal and noise vectors which are both independent of the measurement matrix, recovery is asymptotically guaranteed in this sense. It is not surprising that our average-case framework leads to an improvement over tree-based RIP since the assumption of independence between signal and measurement matrix rules out the practically unlikely case in which one chooses the very worst possible signal for a given measurement matrix. For a comparison of phase transitions derived from both stable point and RIP analysis in the context of IHT and simple sparsity, we refer the reader to [14, Section 6].

**Extension to noise.** Below the same oversampling thresholds, Theorems 3.9 and 3.12 go further than the tree-based RIP analysis in proving convergence of ITP/NITP to a limit point — whose approximation error is asymptotically bounded by some known stability factor multiplied by the noise level  $\sigma$ . Figure 6 plots the noise stability factor  $\xi(\rho)$  for binary trees, for each of the two stepsize schemes considered ( $\kappa = 1.1$  for NITP). For both ITP and NITP, given any value of  $\rho$  for which the stability factors derived in this paper are defined, they are always lower than the corresponding stability factors derived from analysis of IHT based upon the standard RIP [18]; see [23, Section 2.4] for a comparison.

Comparing Figure 6 with Figure 3, we also observe a significant quantitative improvement in the stability

factors for both algorithm variants compared with those achieved by means of tree-based RIP, in the case of binary trees. It should be pointed out that we have obtained improved stability results by imposing additional restrictions upon the noise, namely that the noise is Gaussian distributed and independent of the signal and measurement matrix. This assumption is in keeping with our aim of exploiting average-case assumptions. Our analysis could, however, be altered to deal with the case of non-independent noise by making more use of the RIP, though this would lead to larger stability constants.

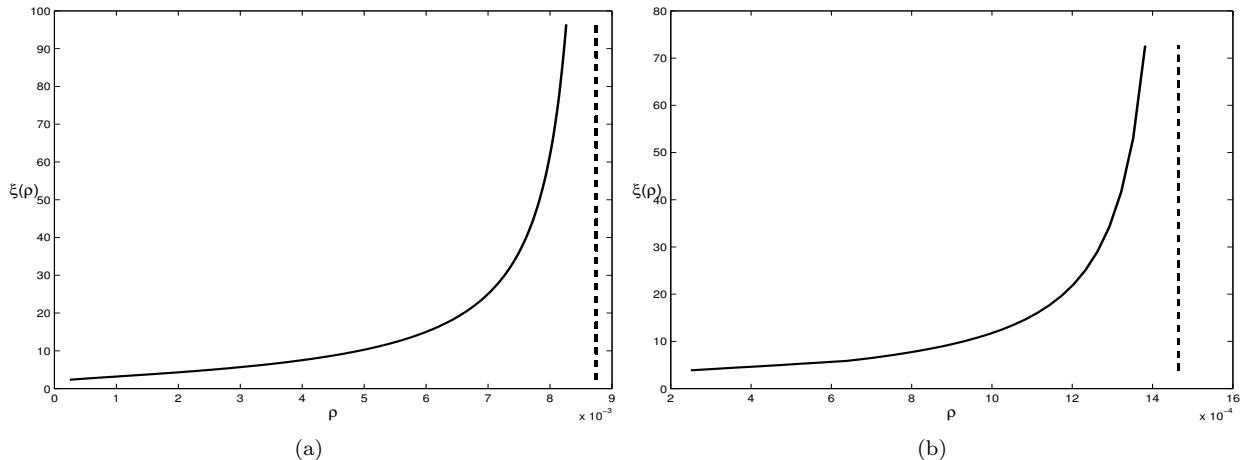


Figure 6: Plot of the stability factor  $\xi(\rho)$  from stable point analysis for binary trees: (a) ITP; (b) NITP.

**Extension to tree compressible signals** While we have assumed so far in this paper that signals are exactly  $k$ -tree sparse, it is more realistic to expect that signals are *tree compressible*, meaning that they are well approximated by a  $k$ -tree sparse vector. An important consideration for any compressed sensing recovery analysis is, therefore, whether it can be extended to the tree compressible case. From the point of view of worst-case analysis, a difference emerges in this respect between standard and tree-based compressed sensing. In the case of standard compressed sensing, the extension to compressible signals can be achieved using the RIP, which can be used to bound the amplification factor of the signal tail [21]. However, it was argued in [1] that the RIP is not sufficient to control this amplification factor for more general structured sparsity models (including the tree-based model). This deficit was partially addressed by the introduction of the Restricted Amplification Property (RAmP), and the extension to model-compressible signals was established provided the sparsity model has a certain ‘nested’ property [1], which unfortunately is not the case for the rooted tree model.

On the other hand, the stable point approach in which we consider independent Gaussian noise is much more amenable to the analysis of the tree-compressible case. In [23, Chapter 7], the main results of the present paper are extended to the tree-compressible case. More precisely, the assumption that  $x^*$  is  $k$ -tree sparse is relaxed, and  $x_k^*$  is defined to be the closest  $k$ -tree sparse approximation to  $x^*$ , namely  $x_k^* := \mathcal{P}_k(x^*)$ . Defining  $\Lambda^k$  to be the support of this optimal tree-sparse approximation, that is  $\Lambda^k := \text{supp}(x_k^*)$ , a measure of unrecoverable energy,  $\Sigma$ , is defined to be

$$\Sigma := \sigma + \|x_{\Lambda^k}^*\|,$$

which represents the combined inaccuracy due to both measurement noise and signal model violation. It is shown in [23, Theorems 7.23 and 7.29] that, beneath the same oversampling thresholds given in Theorems 4.11 and 4.15 of the present paper, the approximation error of the output of ITP/NITP amplifies the unrecoverable energy by no more than some (different) stability factor. See [23, Chapter 7] for an explicit quantification of the stability factor in these cases.

## A Proofs for the tree-based RIP analysis

**Roadmap to the proofs in the Appendices** We begin with a brief roadmap of the proofs found in the following appendices. The current appendix gives proofs for the tree-based RIP analysis, and Appendix B gives proofs for the stable point analysis. In both cases, we first obtain recovery conditions for deterministic matrices (in Sections A.1 and B.1 respectively). We then perform a probabilistic analysis of these conditions



for Gaussian matrices in the simplified proportional-growth asymptotic (in Sections A.2 and B.2 respectively). In both cases, the analysis for Gaussian matrices relies on large deviations results for certain quantities related to Gaussian matrices (including bounds on tree-based RIP constants). These large deviations results, which extend to the tree-based setting those given originally in [3, 14], are stated and proved in Appendix C.

## A.1 Deterministic recovery conditions

The following lemma gives some further consequences of the tree-based RIP.

**Lemma A.1 (Consequences of the tree-based RIP)** *Given some positive integer  $s$ , suppose that  $A \in \mathbb{R}^{n \times N}$  has lower and upper tree-based RIP constants  $TL_s$  and  $TU_s$  respectively, as defined in (3.1). Let  $\Omega \in \mathcal{T}_s$ , and let  $\Omega = \Omega_1 \cup \Omega_2$  where  $|\Omega_1| = s_1$ ,  $|\Omega_2| = s_2$  and  $s = s_1 + s_2$ . Then*

$$\|A_{\Omega}^T y\| \leq \sqrt{1 + U_s} \|y\| \quad \text{for all } y \in \mathbb{R}^n; \quad (\text{A.57})$$

$$(1 - L_s) \|x\| \leq \|A_{\Omega}^T A_{\Omega} x\| \leq (1 + U_s) \|x\| \quad \text{for all } x \in \mathbb{R}^s; \quad (\text{A.58})$$

$$\frac{1}{1 + U_s} \|x\| \leq \|(A_{\Omega}^T A_{\Omega})^{-1} x\| \leq \frac{1}{1 - L_s} \|x\| \quad \text{for all } x \in \mathbb{R}^s; \quad (\text{A.59})$$

$$\|A_{\Omega}^{\dagger} y\| \leq \frac{1}{\sqrt{1 - L_s}} \|y\| \quad \text{for all } y \in \mathbb{R}^n, \text{ provided } A_{\Omega}^{\dagger} \text{ is well-defined}; \quad (\text{A.60})$$

$$\|A_{\Omega_1}^T A_{\Omega_2} z\| \leq \frac{1}{2} (L_s + U_s) \|z\| \quad \text{for all } z \in \mathbb{R}^{s_2}; \quad (\text{A.61})$$

$$\|(I - \omega A_{\Omega}^T A_{\Omega}) x\| \leq \max\{\omega(1 + U_s) - 1, 1 - \omega(1 - L_s)\} \|x\| \quad \text{for all } x \in \mathbb{R}^s \text{ and all } \omega > 0. \quad (\text{A.62})$$

**Proof:** All the above results were proved for the standard notion of RIP in [4, Lemma 15]. The results extend trivially by restricting all support sets to rooted trees.  $\square$

Next, by largely following the analysis in [18], we use the tree-based RIP to obtain a result for generic ITP with bounded stepsize.

**Lemma A.2 (Iteration invariant for bounded stepsize)** *Consider Problem 2. Let the stepsizes of generic ITP satisfy*

$$\underline{\alpha} \leq \alpha^m \leq \bar{\alpha} \quad (\text{A.63})$$

for all  $m \geq 0$ . Then

$$\|x^{m+1} - x^*\| \leq \sqrt{3} \max\{\bar{\alpha}(1 + TU_{3k}) - 1, 1 - \underline{\alpha}(1 - TL_{3k})\} \|x^m - x^*\| + \bar{\alpha} \sqrt{3(1 + TU_{2k})} \|e\|. \quad (\text{A.64})$$

**Proof:** Let us write  $a^m := x^m + \alpha^m A^T (b - Ax^m)$ , which can be rearranged to give

$$a^m = x^m + \alpha^m A^T (Ax^* + e - Ax^m) = x^* + (I - \alpha^m A^T A)(x^m - x^*) + \alpha^m A^T e. \quad (\text{A.65})$$

Let  $\Gamma^m = \text{supp}(x^m)$ , let  $\Gamma^{m+1} = \text{supp}(x^{m+1})$  and let us further define

$$\Omega = \Lambda \cup \Gamma^m \cup \Gamma^{m+1}, \quad (\text{A.66})$$

where  $\Lambda$  is defined in (4.33). By (2.4), we have

$$\|a_{\Lambda}^m\|^2 \leq \|a_{\Gamma^{m+1}}^m\|^2,$$

which cancels to give

$$\|a_{\Lambda \setminus \Gamma^{m+1}}^m\|^2 \leq \|a_{\Gamma^{m+1} \setminus \Lambda}^m\|^2. \quad (\text{A.67})$$

Substituting (A.65) into (A.67) gives

$$\begin{aligned} & \left\| \{x^* + (I - \alpha^m A^T A)(x^m - x^*) + \alpha^m A^T e\}_{\Lambda \setminus \Gamma^{m+1}} \right\| \\ & \leq \left\| \{x^* + (I - \alpha^m A^T A)(x^m - x^*) + \alpha^m A^T e\}_{\Gamma^{m+1} \setminus \Lambda} \right\|, \end{aligned}$$

and the triangle inequality, along with  $x_{\Gamma^{m+1} \setminus \Lambda}^* = 0$ , implies

$$\begin{aligned} & \left\| x_{\Lambda \setminus \Gamma^{m+1}}^* \right\| - \left\| \{(I - \alpha^m A^T A)(x^m - x^*) + \alpha^m A^T e\}_{\Lambda \setminus \Gamma^{m+1}} \right\| \\ & \leq \left\| \{(I - \alpha^m A^T A)(x^m - x^*) + \alpha^m A^T e\}_{\Gamma^{m+1} \setminus \Lambda} \right\|. \end{aligned} \quad (\text{A.68})$$

The sets  $\Lambda \setminus \Gamma^{m+1}$  and  $\Gamma^{m+1} \setminus \Lambda$  are disjoint, and we may therefore apply the Cauchy-Schwarz inequality, namely  $(a+b)^2 \leq \sqrt{2}(a^2+b^2)$ , to (A.68), yielding

$$\|x_{\Lambda \setminus \Gamma^{m+1}}^*\| \leq \sqrt{2} \|\{(I - \alpha^m A^T A)(x^m - x^*) + \alpha^m A^T e\}_{\Lambda \cup \Gamma^{m+1}}\|,$$

from which a further application of the triangle inequality and (A.66) leads us to deduce

$$\|x_{\Lambda \setminus \Gamma^{m+1}}^*\| \leq \sqrt{2} \{\|(I - \alpha^m A_{\Omega}^T A_{\Omega})(x^m - x^*)_{\Omega}\| + \alpha^m \|A_{\Lambda \cup \Gamma^{m+1}}^T e\|\}. \quad (\text{A.69})$$

Meanwhile, splitting on  $\Gamma^{m+1}$  and  $\Lambda \setminus \Gamma^{m+1}$ , and using the definition of  $\Gamma^{m+1} = \text{supp}(x^{m+1})$ ,

$$\begin{aligned} \|x^{m+1} - x^*\|^2 &= \|(x^{m+1} - x^*)_{\Gamma^{m+1}}\|^2 + \|(x^{m+1} - x^*)_{\Lambda \setminus \Gamma^{m+1}}\|^2 \\ &= \|\{(I - \alpha^m A^T A)(x^m - x^*) + \alpha^m A^T e\}_{\Gamma^{m+1}}\|^2 + \|x_{\Lambda \setminus \Gamma^{m+1}}^*\|^2, \end{aligned}$$

where the second inequality follows from (A.65). We then apply the triangle inequality and (A.66) to deduce

$$\begin{aligned} \|x^{m+1} - x^*\|^2 &\leq \|\{(I - \alpha^m A^T A)(x^m - x^*)\}_{\Gamma^{m+1}}\|^2 + \|\{\alpha^m A^T e\}_{\Gamma^{m+1}}\|^2 + \|x_{\Lambda \setminus \Gamma^{m+1}}^*\|^2 \\ &\leq \|(I - \alpha^m A_{\Omega}^T A_{\Omega})(x^m - x^*)_{\Omega}\|^2 + \alpha^m \|A_{\Lambda \cup \Gamma^{m+1}}^T e\|^2 + \|x_{\Lambda \setminus \Gamma^{m+1}}^*\|^2. \end{aligned} \quad (\text{A.70})$$

Substituting (A.69) into (A.70) then gives

$$\|x^{m+1} - x^*\|^2 \leq 3 \{\|(I - \alpha^m A_{\Omega}^T A_{\Omega})(x^m - x^*)_{\Omega}\|^2 + \alpha^m \|A_{\Lambda \cup \Gamma^{m+1}}^T e\|^2\}. \quad (\text{A.71})$$

Since  $|\Omega| \leq 3k$  and  $|\Lambda \cup \Gamma^{m+1}| \leq 2k$ , the result now follows by applying (A.57), (A.62) and (A.63) to (A.71), and taking square roots.  $\square$

Both the ITP and NITP stepsize schemes have bounded stepsizes; this is trivial in the case of ITP, while bounds for NITP are given next.

**Lemma A.3 (NITP stepsize bounds)** *Let  $\alpha^m$  be chosen according to Algorithm 2.2. Then*

$$\frac{1}{\kappa(1 + TU_{2k})} \leq \alpha^m \leq \frac{1}{1 - TL_k}. \quad (\text{A.72})$$

**Proof of Lemma A.3:** If (2.7) is accepted, then  $\alpha^m \leq 1/(1 - TL_k)$  by (3.8). On the other hand, if (2.7) is rejected, the backtracking phase can only reduce the stepsize further, which proves the upper bound in (A.72). To prove the lower bound, we also distinguish two cases. If (2.7) is accepted, then  $\alpha^m \geq 1/(1 + TU_k)$  by (3.8). Since  $\kappa > 1$ , and since  $TU_{2k} \geq TU_k$  by the nonincreasing property of tree-based RIP constants, the lower bound in (A.72) holds in this case. On the other hand, if (2.7) is rejected, the penultimate stepsize calculated in the backtracking phase must also have been rejected. Writing  $\tilde{\alpha}^m$  for the penultimate stepsize, since  $\tilde{\alpha}^m$  was rejected, we have

$$\tilde{\alpha}^m \geq (1 - c) \frac{\|\tilde{x}^{m+1} - x^m\|^2}{\|A(\tilde{x}^{m+1} - x^m)\|^2} \geq \frac{1 - c}{1 + TU_{2k}}, \quad (\text{A.73})$$

where the last step follows from (3.8). But  $\alpha^m = \tilde{\alpha}^m / [\kappa(1 - c)]$ , which combines with (A.73) to give the lower bound in (A.72) in this case also.  $\square$

We may therefore deduce the following results.

**Theorem A.4 (Iteration invariant for ITP)** *Consider Problem 2. Then the iterates of ITP with stepsize  $\alpha$  satisfy*

$$\|x^{m+1} - x^*\| \leq \mu^{ITP_{\alpha}} \|x^m - x^*\| + \xi^{ITP_{\alpha}} \|e\|, \quad (\text{A.74})$$

where  $\mu^{ITP_{\alpha}}$  and  $\xi^{ITP_{\alpha}}$  are defined in (3.9) and (3.10), respectively.

**Proof:** For ITP with stepsize  $\alpha$ , we have  $\underline{\alpha} = \bar{\alpha} = \alpha$ , and the result follows by applying Lemma A.2.  $\square$

**Theorem A.5 (Iteration invariant for NITP)** Consider Problem 2 and suppose Assumption 2 holds. Then the iterates of NITP with shrinkage parameter  $\kappa$  satisfy

$$\|x^{m+1} - x^*\| \leq \mu^{NITP_\kappa} \|x^m - x^*\| + \xi^{NITP_\kappa} \|e\|, \quad (\text{A.75})$$

where

$$\mu^{NITP_\kappa} := \sqrt{3} \max \left\{ \frac{1 + TU_{3k}}{1 - TL_k} - 1, 1 - \frac{1 - TL_{3k}}{\kappa(1 + TU_{2k})} \right\} \quad (\text{A.76})$$

and

$$\xi^{NITP_\kappa} := \frac{\sqrt{3(1 + TU_{2k})}}{1 - TL_k}. \quad (\text{A.77})$$

**Proof:** For a given  $\kappa > 1$ , the stepsize bounds (A.3) apply to NITP, and the result follows by applying Lemma A.2 with  $\underline{\alpha} := 1/(1 - L_k)$  and  $\bar{\alpha} := 1/[\kappa(1 + U_{2k})]$ .  $\square$

In order to prove recovery results, we will need the following lemma.

**Lemma A.6** Suppose there exist  $\mu \in [0, 1)$  and  $\xi > 0$  such that the sequence of iterates  $\{x^m\}$  satisfies, for each  $m \geq 0$ ,

$$\|x^{m+1} - x^*\| \leq \mu \|x^m - x^*\| + \xi \|e\|. \quad (\text{A.78})$$

Then, for all  $m \geq 0$ ,

$$\|\bar{x} - x^*\| \leq \mu^m \|x^*\| + \frac{\xi}{1 - \mu} \|e\|. \quad (\text{A.79})$$

**Proof:** We first prove by induction that, for all  $m \geq 0$ ,

$$\|x^m - x^*\| \leq \mu^m \|x^*\| + \xi \left( \frac{1 - \mu^m}{1 - \mu} \right) \|e\|. \quad (\text{A.80})$$

Supposing (A.80) holds for some  $m \geq 0$ , then we may apply (A.78) to (A.80) to deduce

$$\begin{aligned} \|x^{m+1} - x^*\| &\leq \mu \left[ \mu^m \|x^*\| + \xi \left( \frac{1 - \mu^m}{1 - \mu} \right) \|e\| \right] + \xi \|e\| \\ &= \mu^{m+1} \|x^*\| + \left[ \xi \left( 1 + \frac{\mu - \mu^{m+1}}{1 - \mu} \right) \right] \|e\| \\ &= \mu^{m+1} \|x^*\| + \xi \left( \frac{1 - \mu^{m+1}}{1 - \mu} \right) \|e\|, \end{aligned}$$

and so (A.80) also holds for  $m + 1$ . Since  $x^0 = 0$ , the result holds trivially for  $m = 0$ , and therefore for all  $m \geq 0$  by induction. Since  $\mu^m \in (0, 1)$  for all  $m \geq 0$ , (A.79) now follows.  $\square$

Provided  $\mu < 1$ , the  $\mu^m \|x^*\|$  term in (A.79) tends to zero, and the expression  $\xi/(1 - \mu)$  may be viewed as a *stability factor*, giving a limiting bound on the approximation error as a multiple of the noise level  $\|e\|$ . We now proceed to the proof of the recovery results for ITP and arbitrary matrices.

**Proof of Theorem 3.3:** The result follows by combining Theorem A.4 and Lemma A.6.  $\square$

The result for NITP and arbitrary matrices follows.

**Theorem A.7 (Deterministic recovery result for NITP)** Consider Problem 2. Let  $\mu^{NITP_\kappa}$  and  $\xi^{NITP_\kappa}$  be defined as in Theorem A.5. Then, provided  $\mu^{NITP_\kappa} < 1$ , the output,  $\hat{x}$ , of NITP with shrinkage parameter  $\kappa$  at iteration  $m$ , satisfies

$$\|\hat{x} - x^*\| \leq (\mu^{NITP_\kappa})^m \|x^*\| + \frac{\xi^{NITP_\kappa}}{1 - \mu^{NITP_\kappa}} \|e\|. \quad (\text{A.81})$$

**Proof:** The result follows by combining Theorem A.5 and Lemma A.6.  $\square$

Though both Theorems 3.3 and A.7 give a limiting bound on the approximation error, they do not necessarily imply convergence of the algorithm. In the simplified noiseless case however, both results can be used to deduce convergence to  $x^*$  at a linear rate.

## A.2 Analysis for Gaussian matrices

The next two lemmas are needed to enable a translation of Theorems 3.3 and A.7 for arbitrary matrices into the asymptotic framework for Gaussian matrices.

**Lemma A.8** *For some  $\tau < 1$ , define the set  $\mathcal{Z} := (0, \tau)^p \times (0, \infty)^q$  and let  $F : \mathcal{Z} \rightarrow \mathbb{R}$  be continuously differentiable on  $\mathcal{Z}$ . Let  $A \in \mathbb{R}^{n \times N}$  be a Gaussian matrix with tree-based RIP constants  $TL_k, \dots, TL_{pk}$  and  $TU_k, \dots, TU_{qk}$ , and let  $\mathcal{TL}(\rho), \dots, \mathcal{TL}(p\rho)$  and  $\mathcal{TU}(\rho), \dots, \mathcal{TU}(q\rho)$  be defined as in Definition 3.6. Define  $\mathbf{1}$  to be the vector of all ones, and*

$$\begin{aligned} z(k, n, N) &:= [TL_k, \dots, TL_{pk}, TU_k, \dots, TU_{qk}], \\ z(\rho) &:= [\mathcal{TL}(\rho), \dots, \mathcal{TL}(p\rho), \mathcal{TU}(\rho), \dots, \mathcal{TU}(q\rho)]. \end{aligned}$$

*Suppose, for all  $t \in \mathcal{Z}$ ,  $(\nabla F[t])_i \geq 0$  for all  $i = 1, \dots, p + q$  and there exists  $j \in \{1, \dots, p\}$  such that  $(\nabla F[t])_j > 0$ . Then, for any  $\epsilon \in (0, 1)$ , in the simplified proportional-growth asymptotic,*

$$\mathbb{P}(F[z(k, n, N)] < F[z((1 + \epsilon)\rho)]) \rightarrow 1 \quad \text{as } n \rightarrow \infty, \quad (\text{A.82})$$

*exponentially in  $n$  on the draw of  $A$ . Also,  $F[z(\rho)]$  is strictly increasing in  $\rho$ .*

**Proof:** A proof was given in [5, Lemma 12] for the case where  $\mathcal{TL}(\cdot)$  and  $\mathcal{TU}(\cdot)$  are replaced by  $\mathcal{L}(\delta, \cdot)$  and  $\mathcal{U}(\delta, \cdot)$ , bounds on the standard notion of RIP constants given in [3]. Note first that a function that depends only upon  $\rho$  is a trivial special case of a function that depends upon both  $\delta$  and  $\rho$ . Only two assumptions are made in the proof concerning the bounds: first that they are indeed upper bounds, and second that  $\mathcal{L}(\delta, \rho)$  is strictly increasing in  $\rho$  and  $\mathcal{U}(\delta, \rho)$  is nondecreasing in  $\rho$ . The first condition holds in our case by Theorem 3.7, and it is straightforward to show that the second property also holds in our case. More precisely,  $\mathcal{TL}(\rho)$  and  $\mathcal{TU}(\rho)$  are both strictly increasing on  $\rho \in (0, 1)$ . It follows that the argument in [5, Lemma 12] extends.  $\square$

**Lemma A.9** *For some  $\tau < 1$ , define the set  $\mathcal{Z} := (0, \tau)^p \times (0, \infty)^q$  and let  $F, G, H : \mathcal{Z} \rightarrow \mathbb{R}$  satisfy the conditions of Lemma A.8. Suppose that*

$$\mu(k, n, N) = \max \{F[z(k, n, N)], G[z(k, n, N)]\}, \quad \xi(k, n, N) = H[z(k, n, N)], \quad (\text{A.83})$$

*and*

$$\mu(\rho) = \max \{F[z(\rho)], G[z(\rho)]\}, \quad \xi(\rho) = H[z(\rho)]. \quad (\text{A.84})$$

*Then  $\mu(\rho)$  and  $\xi(\rho)$  are both strictly increasing in  $\rho$  and, for any  $\epsilon \in (0, 1)$ , in the proportional-growth asymptotic,*

$$\mathbb{P}\{\mu(k, n, N) \geq \mu((1 + \epsilon)\rho)\} \rightarrow 0, \quad (\text{A.85})$$

*and*

$$\mathbb{P}\{\xi(k, n, N) \geq \xi((1 + \epsilon)\rho)\} \rightarrow 0, \quad (\text{A.86})$$

*both exponentially in  $n$ . Furthermore, define  $\hat{\rho}$  as the unique solution to  $\mu(\rho) = 1$ , and suppose that*

$$\rho < (1 - \epsilon)\hat{\rho}. \quad (\text{A.87})$$

*Then*

$$\mu((1 + \epsilon)\rho) < 1, \quad (\text{A.88})$$

*and, in the simplified proportional-growth asymptotic,*

$$\mathbb{P}\{\mu(k, n, N) \geq 1\} \rightarrow 0, \quad (\text{A.89})$$

*exponentially in  $n$ .*

**Proof:** By assumption, we may apply Lemma A.8 to each of  $F(z)$ ,  $G(z)$  and  $H(z)$ , deducing from (A.82) that

$$\mathbb{P}(F[z(k, n, N)] < F[z((1 + \epsilon)\rho)]) \rightarrow 1 \quad \text{as } n \rightarrow \infty, \quad (\text{A.90})$$

$$\mathbb{P}(G[z(k, n, N)] < G[z((1 + \epsilon)\rho)]) \rightarrow 1 \quad \text{as } n \rightarrow \infty, \quad (\text{A.91})$$

$$\mathbb{P}(H[z(k, n, N)] < H[z((1 + \epsilon)\rho)]) \rightarrow 1 \quad \text{as } n \rightarrow \infty, \quad (\text{A.92})$$

exponentially in  $n$ , and that  $F[z(\rho)]$ ,  $G[z(\rho)]$  and  $H[z(\rho)]$  are each strictly increasing in  $\rho$ , from which it immediately follows that both  $\mu(\rho)$  and  $\xi(\rho)$  are also strictly increasing in  $\rho$ . Combining (A.83), (A.84), (A.91) and (A.92), we have

$$\begin{aligned} & \mathbb{P}\{\mu(k, n, N) \geq \mu((1 + \epsilon)\rho)\} \\ = & \mathbb{P}\left\{\max\{F[z(k, n, N)], G[z(k, n, N)]\} \geq \max\{F[z((1 + \epsilon)\rho)], G[z((1 + \epsilon)\rho)]\}\right\} \\ \leq & \mathbb{P}\{F[z(k, n, N)] \geq F[z((1 + \epsilon)\rho)]\} + \mathbb{P}\{G[z(k, n, N)] \geq G[z((1 + \epsilon)\rho)]\} \\ \rightarrow & 0 \text{ as } n \rightarrow \infty, \end{aligned} \tag{A.93}$$

and therefore (A.85) holds. Meanwhile, combining (A.83), (A.84) and (A.92) immediately yields (A.86). Now suppose (A.87) holds. Since  $1 - \epsilon < (1 + \epsilon)^{-1}$  for any  $\epsilon \in (0, 1)$ , (A.87) implies that

$$(1 + \epsilon)\rho < \hat{\rho}, \tag{A.94}$$

Since  $\mu(\rho)$  is strictly increasing in  $\rho$ , it follows from (A.94) and the definition of  $\hat{\rho}$  that

$$\mu((1 + \epsilon)\rho) < \mu(\hat{\rho}) = 1,$$

which proves (A.88), and from which it also follows that

$$\mathbb{P}\{\mu(k, n, N) \geq 1\} \leq \mathbb{P}\{\mu(k, n, N) \geq \mu((1 + \epsilon)\rho)\},$$

to which we may apply (A.93) to deduce (A.89).  $\square$

We now proceed to the proofs of the main results.

**Proof of Theorem 3.9:** Select  $\epsilon \in (0, 1)$ , fix  $\tau < 1$  and let

$$z(k, n, N) := [TL_{3k}, TU_{2k}, TU_{3k}] \quad \text{and} \quad z(\rho) := [\mathcal{TL}(3\rho), \mathcal{TU}(2\rho), \mathcal{TU}(3\rho)].$$

Define  $\mathcal{Z} := (0, \tau) \times (0, \infty)^2$ , and define the functions  $F_\alpha(z), G_\alpha(z), H_\alpha(z) : \mathcal{Z} \rightarrow \mathbb{R}$  as

$$F_\alpha(z) = F_\alpha(z_1, z_2, z_3) := \sqrt{3}[\alpha(1 + z_3) - 1], \tag{A.95}$$

$$G_\alpha(z) = G_\alpha(z_1, z_2, z_3) := \sqrt{3}[1 - \alpha(1 - z_1)], \tag{A.96}$$

$$H_\alpha(z) = H_\alpha(z_1, z_2, z_3) := \alpha\sqrt{3(1 + z_2)}, \tag{A.97}$$

noting that

$$\mu^{ITP_\alpha} = \max\{F_\alpha[z(k, n, N)], G_\alpha[z(k, n, N)]\}, \quad \xi^{ITP_\alpha} = H_\alpha[z(k, n, N)],$$

where  $\mu^{ITP_\alpha}$  and  $\xi^{ITP_\alpha}$  are defined in (3.9) and (3.10) respectively, and

$$\mu^{ITP_\alpha}(\rho) = \max\{F_\alpha[z(\rho)], G_\alpha[z(\rho)]\}, \quad \xi^{ITP_\alpha}(\rho) = H_\alpha[z(\rho)],$$

where  $\mu^{ITP_\alpha}(\rho)$  and  $\xi^{ITP_\alpha}(\rho)$  are defined in (3.20) and (3.21) respectively. Now  $F_\alpha(z)$ ,  $G_\alpha(z)$  and  $H_\alpha(z)$  are continuously differentiable and nondecreasing in  $(z_1, z_2, z_3) \in \mathcal{Z}$ , and strictly increasing in  $z_3$ ,  $z_1$  and  $z_2$  respectively due to  $\alpha > 0$ , and therefore each satisfies the conditions of Lemma A.8. We may therefore apply Lemma A.9, deducing

$$\mathbb{P}\{\mu^{ITP_\alpha} \geq \mu_{RIP}^{ITP_\alpha}((1 + \epsilon)\rho)\} \rightarrow 0, \tag{A.98}$$

and

$$\mathbb{P}\{\xi^{ITP_\alpha} \geq \xi_{RIP}^{ITP_\alpha}((1 + \epsilon)\rho)\} \rightarrow 0, \tag{A.99}$$

exponentially in  $n$ , and furthermore that  $\mu_{RIP}^{ITP_\alpha}(\rho)$  and  $\xi_{RIP}^{ITP_\alpha}(\rho)$  are both strictly increasing in  $\rho$ , from which it follows that  $\hat{\rho}_{RIP}^{ITP_\alpha}$  is unique. Since (3.22) holds, we may also use Lemma A.9 to deduce (3.23), and furthermore that

$$\mathbb{P}\{\mu^{ITP_\alpha} \geq 1\} \rightarrow 0, \tag{A.100}$$

exponentially in  $n$ , and we may apply Corollary 3.3 to deduce (3.11) with probability tending to 1 exponentially in  $n$ . Since  $\mu_{RIP}^{ITP_\alpha}(\rho)$  and  $\xi_{RIP}^{ITP_\alpha}(\rho)$  are strictly increasing in  $\rho$ , (3.24) now follows from (3.11), (A.98) and (A.99).  $\square$

**Proof of Corollary 3.10:** Since we consider Problem 1, we have  $e := 0$ . Provided (3.22) holds, we can apply Theorem 3.9 with  $e := 0$ , deducing that, for any  $m \geq 0$ ,

$$\|x^m - x^*\| \leq \left( \mu_{RIP}^{ITP\alpha}(\delta, (1 + \epsilon)\rho) \right)^m \|x^*\|,$$

where

$$\mu_{RIP}^{ITP\alpha}((1 + \epsilon)\rho) < 1,$$

and so we have convergence to  $x^*$  with convergence rate  $\mu_{RIP}^{ITP\alpha}((1 + \epsilon)\rho)$ .  $\square$

**Proof of Theorem 3.12:** Select  $\epsilon \in (0, 1)$ , fix  $\tau < 1$  and let

$$z(k, n, N) := [TL_k, TL_{3k}, TU_{2k}, TU_{3k}] \quad \text{and} \quad z(\rho) := [\mathcal{TL}(\rho), \mathcal{TL}(3\rho), \mathcal{TU}(2\rho), \mathcal{TU}(3\rho)].$$

Define  $\mathcal{Z} := (0, \tau)^2 \times (0, \infty)^2$ , and define the functions  $F_\alpha(z), G_\alpha(z), H_\alpha(z) : \mathcal{Z} \rightarrow \mathbb{R}$  as

$$\begin{aligned} F_\kappa(z) &= F_\kappa(z_1, z_2, z_3, z_4) := \sqrt{3} \left[ \frac{1 + z_4}{1 - z_1} - 1 \right], \\ G_\kappa(z) &= G_\kappa(z_1, z_2, z_3, z_4) := \sqrt{3} \left[ 1 - \frac{1 - z_2}{\kappa(1 + z_3)} \right], \\ H_\kappa(z) &= H_\kappa(z_1, z_2, z_3, z_4) := \frac{\sqrt{3(1 + z_3)}}{1 - z_1}, \end{aligned}$$

noting that

$$\mu^{NITP\kappa} = \max \{ F_\kappa[z(k, n, N)], G_\kappa[z(k, n, N)] \}, \quad \xi^{NITP\kappa} = H_\kappa[z(k, n, N)],$$

where  $\mu^{NITP\kappa}$  and  $\xi^{NITP\kappa}$  are defined in (A.76) and (A.77) respectively, and

$$\mu_{RIP}^{NITP\kappa}(\rho) = \max \{ F_\kappa[z(\delta, \rho)], G_\kappa[z(\rho)] \}, \quad \xi_{RIP}^{NITP\kappa}(\rho) = H_\kappa[z(\rho)],$$

where  $\mu_{RIP}^{NITP\kappa}(\rho)$  and  $\xi_{RIP}^{NITP\kappa}(\rho)$  are defined in (3.26) and (3.27) respectively. Now  $F_\kappa(z), G_\kappa(z)$  and  $H_\kappa(z)$  are continuously differentiable and nondecreasing in  $(z_1, z_2, z_3, z_4)$ , and strictly increasing componentwise in  $(z_1, z_4), (z_2, z_3)$  and  $(z_1, z_3)$  respectively, and therefore each satisfies the conditions of Lemma A.8. We may therefore apply Lemma A.9, deducing

$$\mathbb{P}\{ \mu^{NITP\kappa} \geq \mu_{RIP}^{NITP\kappa}((1 + \epsilon)\rho) \} \rightarrow 0 \tag{A.101}$$

and

$$\mathbb{P}\{ \xi^{NITP\kappa} \geq \xi_{RIP}^{NITP\kappa}((1 + \epsilon)\rho) \} \rightarrow 0, \tag{A.102}$$

exponentially in  $n$ , and furthermore that  $\mu_{RIP}^{NITP\kappa}(\rho)$  and  $\xi_{RIP}^{NITP\kappa}(\rho)$  are both strictly increasing in  $\rho$ , from which it follows that  $\hat{\rho}_{RIP}^{NITP\kappa}$  is unique. Since (3.28) holds, we may also use Lemma A.9 to deduce (3.29), and furthermore that

$$\mathbb{P}\{ \mu^{NITP\kappa} \geq 1 \} \rightarrow 0, \tag{A.103}$$

exponentially in  $n$ , and we may apply Corollary A.7 to deduce (A.81) with probability tending to 1 exponentially in  $n$ . Since  $\mu_{RIP}^{NITP\kappa}(\rho)$  and  $\xi_{RIP}^{NITP\kappa}(\rho)$  are strictly increasing in  $\rho$ , (3.30) now follows from (A.81), (A.101) and (A.102).  $\square$

**Proof of Corollary 3.13:** Since we consider Problem 1, we have  $e := 0$ . Provided (3.28) holds, we can apply Theorem 3.12 with  $e := 0$ , deducing that, for any  $m \geq 0$ ,

$$\|x^m - x^*\| \leq \left( \mu_{RIP}^{NITP\kappa}((1 + \epsilon)\rho) \right)^m \|x^*\|,$$

where

$$\mu_{RIP}^{NITP\kappa}((1 + \epsilon)\rho) < 1,$$

and so we have convergence to  $x^*$  with convergence rate  $\mu_{RIP}^{NITP\kappa}((1 + \epsilon)\rho)$ .  $\square$

## B Proofs for the tree-based stable point analysis

### B.1 Analysis for deterministic matrices

We will follow the approach first introduced by the present authors in [14], central to which is the concept of an  $\underline{\alpha}$ -stable point, defined in Definition 4.1.

We will analyse the stable points of generic ITP, and our final goal is to prove quantitative conditions that guarantee that all stable points of the algorithm are ‘close’ to the original signal  $x$ , in the context of Gaussian matrices. Provided we also have guaranteed convergence to some stable point, we may then conclude that ITP outputs a good approximation to  $x$ . For this reason, the results derived in this section come in two parts: a necessary condition for there to be a stable point on some support  $\Gamma$ , and conditions guaranteeing convergence to some stable point for our two stepsize schemes.

#### B.1.1 A necessary condition for the existence of a stable point

Any  $\underline{\alpha}$ -stable point of generic ITP may also be characterized as a minimum-norm solution on some  $k$ -subspace.

**Lemma B.1** *Suppose Assumption 2 holds and suppose  $\bar{x}$  is an  $\underline{\alpha}$ -stable point of generic ITP on  $\Gamma$  for some  $\underline{\alpha} > 0$ . Then  $\bar{x}_\Gamma = A_\Gamma^\dagger b$ .*

**Proof:** It follows from (4.31) that  $A_\Gamma^T(b - A_\Gamma \bar{x}_\Gamma) = 0$  where  $\text{supp}(\bar{x}) \subseteq \Gamma$  and  $|\Gamma| = k$ . Under Assumption 2, the pseudoinverse  $A_\Gamma^\dagger$  is well-defined and we may rearrange to give  $\bar{x}_\Gamma = A_\Gamma^\dagger b$ .  $\square$

While this lemma tells us that any stable point is necessarily a minimum-norm solution on some  $k$ -subspace, the converse may not hold. We next prove Theorem 4.2, which gives a necessary condition for a stable point on a given support.

**Proof of Theorem 4.2:** Supposing that  $\bar{x}$  is an  $\underline{\alpha}$ -stable point on  $\Gamma$ , choosing  $\Omega := \Lambda$  in (4.32) yields

$$\|\bar{x}_{\Gamma \setminus \Lambda}\|^2 \geq \underline{\alpha}^2 \|A_{\Lambda \setminus \Gamma}^T(b - A\bar{x})\|^2.$$

We may now follow the argument of [14, Theorem 3.2] to deduce (4.34).  $\square$

#### B.1.2 Conditions guaranteeing convergence

In addition to the result of the previous section, in order to show recovery of  $x^*$ , we must also show that ITP converges to an  $\underline{\alpha}$ -stable point. In this section we derive convergence conditions for generic ITP used in conjunction with the two stepsize schemes introduced in Section 2.2. We now introduce some more notation. Recalling (2.2), we let

$$g^m := \nabla \Psi(x^m) \quad \text{and} \quad \Gamma^m := \text{supp}(x^m) \quad \text{for all } m \geq 0, \quad (\text{B.104})$$

where  $\{x^m\}$  are the iterates of generic ITP. Some useful properties of the iterates of ITP are given in the next lemma. Since (2.4) is a Euclidean projection, the argument given in the proof of [14, Lemma 3.4] remains valid.

**Lemma B.2** ([14, Lemma 3.4]) *The iterates of generic ITP satisfy, for all  $m \geq 0$ ,*

$$\|x^{m+1} - x^m\|^2 + 2\alpha^m (g^m)^T (x^{m+1} - x^m) \leq 0 \quad (\text{B.105})$$

and

$$\Psi(x^{m+1}) - \Psi(x^m) = (g^m)^T (x^{m+1} - x^m) + \frac{1}{2} \|A(x^{m+1} - x^m)\|^2. \quad (\text{B.106})$$

A sufficient condition for convergence of generic ITP is given next.

**Lemma B.3 (Sufficient condition for convergence)** *Consider Problem 2. Suppose Assumption 2 holds, and suppose the iterates of generic ITP satisfy*

$$\|x^{m+1} - x^m\|^2 \leq d [\Psi(x^m) - \Psi(x^{m+1})] \quad \text{for all } m \geq 0, \quad (\text{B.107})$$

for some  $d > 0$  which does not depend upon  $m$ , where  $\Psi(\cdot)$  is defined in (2.2). Assume that there exist  $\bar{\alpha} \geq \underline{\alpha} > 0$  such that

$$\bar{\alpha} \geq \alpha^m \geq \underline{\alpha} \quad \text{for all } m \geq 0. \quad (\text{B.108})$$

Then  $x^m \rightarrow \bar{x}$  as  $m \rightarrow \infty$ , where  $\bar{x}$  is an  $\underline{\alpha}$ -stable point of generic ITP.

**Proof:** We may follow the proof of [14, Lemma 3.5] to deduce that  $x^m \rightarrow \bar{x}$ , where  $\bar{x}_\Gamma = A_\Gamma^\dagger b$  and  $\bar{x}_{\Gamma^c} = 0$ , for some  $\Gamma$  such that  $|\Gamma| = k$ . The proof still holds since all that is assumed about the hard threshold projection  $\mathcal{H}_k(\cdot)$  there is that it preserves the value of selected coefficients, a property which is also shared by the tree projection  $\mathcal{P}_k(\cdot)$  by (2.5). Since  $\Gamma = \Gamma^m$  for some  $m \geq 0$ , it follows that, in the case of ITP,  $\Gamma \in \mathcal{T}_k$ . Therefore (4.31) holds for  $\bar{x}$ .

It remains to establish that  $\bar{x}$  satisfies (4.32). Defining

$$\Gamma_1 = \{i \in \Gamma : \bar{x}_i \neq 0\}, \quad (\text{B.109})$$

it follows that  $\Gamma_1 \subseteq \Gamma^m$  for all  $m$  sufficiently large. It follows from (2.5) that, for any  $\Omega \in \mathcal{T}_k$ ,

$$\|x_{\Gamma_1}^{m+1}\|^2 \geq \|\{x^m - \alpha^m g^m\}_\Omega\|^2, \text{ for all } m \geq 0.$$

and therefore, for all  $m$  sufficiently large,

$$\|x_{\Gamma_1}^{m+1}\|^2 + \|x_{\Gamma^{m+1} \setminus \Gamma_1}^{m+1}\|^2 \geq \|x_{\Omega \cap \Gamma_1}^{m+1}\|^2 + \|\{x^m - \alpha^m g^m\}_{\Omega \setminus \Gamma_1}\|^2,$$

which cancels to

$$\|x_{\Gamma_1 \setminus \Omega}^{m+1}\|^2 + \|x_{\Gamma^{m+1} \setminus \Gamma_1}^{m+1}\|^2 \geq \|\{x^m + \alpha^m g^m\}_{\Omega \setminus \Gamma_1}\|^2. \quad (\text{B.110})$$

Furthermore, it follows from (B.109) that

$$\|x_{\Gamma^{m+1} \setminus \Gamma_1}^{m+1}\|^2 \rightarrow 0. \quad (\text{B.111})$$

By (B.108), there exists a convergent subsequence of stepsizes,

$$\alpha^{m_r} \rightarrow \tilde{\alpha} \geq \underline{\alpha} \text{ as } r \rightarrow \infty \quad (\text{B.112})$$

Passing to the limit in (B.110) on the subsequence  $m_r$  for which (B.112) holds, we deduce that  $\|\bar{x}_{\Gamma_1 \setminus \Omega}\| \geq \underline{\alpha} \|\{A^T(b - A\bar{x})\}_{\Omega \setminus \Gamma_1}\|$ , from which it follows trivially that

$$\|\bar{x}_{\Gamma \setminus \Omega}\| \geq \underline{\alpha} \|\{A^T(b - A\bar{x})\}_{\Omega \setminus \Gamma}\|. \quad (\text{B.113})$$

Since (B.113) holds for any  $\Omega \in \mathcal{T}_k$ ,  $\bar{x}$  satisfies (4.32), and the result is proved.  $\square$

**Proof of Theorem 4.3 (ITP convergence):** We mimic the proof of [14, Theorem 3.6]. Let  $m \geq 0$ . Since the support size of the change to the iterates  $x^{m+1} - x^m$  is at most  $2k$ , and since a union of two rooted trees is also a rooted tree, (3.8) with  $s = 2k$  provides  $\|A(x^{m+1} - x^m)\|^2 \leq (1 + TU_{2k})\|x^{m+1} - x^m\|^2$ . Using this bound, and (B.105) with the choice (4.35), in (B.106), we obtain

$$\begin{aligned} \Psi(x^{m+1}) - \Psi(x^m) &\leq -\frac{1}{2\alpha}\|x^{m+1} - x^m\|^2 + \frac{1}{2}(1 + TU_{2k})\|x^{m+1} - x^m\|^2 \\ &= \frac{\alpha(1 + TU_{2k}) - 1}{2\alpha}\|x^{m+1} - x^m\|^2, \end{aligned}$$

which, due to (4.35), implies that (B.107) holds with  $d := 2\alpha/[1 - \alpha(1 + TU_{2k})]$ . Due to (4.35), (B.108) trivially holds with  $\bar{\alpha} = \underline{\alpha} = \alpha$ . Thus Lemma B.3 applies, and the ITP iterates  $x^m$  converge to an  $\alpha$ -stable point.  $\square$

We next obtain a convergence result for NITP. In this case, there is no explicit requirement for a tree-based RIP condition to be satisfied; however, tree-based RIP this time appears in the choice of  $\underline{\alpha}$ .

**Theorem B.4 (NITP convergence)** *Suppose Assumption 2 holds. Then NITP with shrinkage parameter  $\kappa$  converges to a  $[\kappa(1 + TU_{2k})]^{-1}$ -stable point  $\bar{x}$  of generic ITP.*

**Proof:** By replacing  $L_{2k}$  with  $TL_{2k}$ , the proof given for [14, Theorem 3.7] holds.  $\square$

## B.2 Analysis for Gaussian matrices

In this section, we build upon the results for arbitrary matrices in Section B.1 and obtain quantitative oversampling thresholds for ITP algorithms of the form  $\rho < \hat{\rho}$  in the case of Gaussian measurement matrices.



### B.2.1 Proof of ITP recovery results

The present analysis broadly follows the same lines as that in [14], but differs in two respects. First, we switch to using the tree-based tail bounds defined in Section C. Second, since there is now no dependence upon  $\delta$ , we can prove results in the simplified proportional-growth asymptotic (Definition 1.1). The changes are nontrivial, and therefore we present full proofs of the new results. We begin by defining a support set partition.

**Definition B.5 (Support set partition for ITP)** Consider Problem 2 and suppose  $\rho \in (0, 1/2]$  and  $\alpha > 0$ . Given  $\zeta > 0$ , let us write

$$a^*(\rho; \zeta) := a(\rho) + \zeta, \quad (\text{B.114})$$

let us write  $\{\Gamma_i : i \in \mathcal{T}_k\}$  for the set of all possible support sets which form a rooted tree of cardinality  $k$ , and let us disjointly partition  $\mathcal{T}_k := \Theta_n^1 \cup \Theta_n^2$  such that

$$\Theta_n^1 := \left\{ i \in \mathcal{T}_k : \|x_{\Lambda \setminus \Gamma_i}^*\| > \sigma \cdot a^*(\rho; \zeta) \right\} \quad \text{and} \quad \Theta_n^2 := \left\{ i \in \mathcal{T}_k : \|x_{\Lambda \setminus \Gamma_i}^*\| \leq \sigma \cdot a^*(\rho; \zeta) \right\}, \quad (\text{B.115})$$

where  $\Lambda$  is defined in (4.33).

The partition in (B.115) has been defined in such a way that, provided (4.47) holds, an analysis of the stable point condition (4.34) shows that ITP must necessarily converge to some  $\alpha$ -stable point on  $\Gamma_i$  such that  $i \in \Theta_n^2$ , and this is proved in Lemma B.6. On the other hand, it is also possible to use the large deviations results of Section C to bound the error in approximating  $x^*$  by any  $\alpha$ -stable point on  $\Gamma_i$  such that  $i \in \Theta_n^2$ , which is achieved by Lemma B.7. It follows that, for any  $\alpha > 0$ , all  $\alpha$ -stable points have bounded approximation error. Combining these two results, we have convergence to some  $\alpha$ -stable point with guaranteed approximation error, provided the conditions in each lemma hold; combining the conditions leads to the oversampling threshold defined in (4.46).

We first show that, asymptotically, there are no  $\alpha$ -stable points on any  $\Gamma_i$  such that  $i \in \Theta_n^1$ , and we write  $NSP_\alpha$  for this event.

**Lemma B.6** Consider Problem 2 and choose  $\zeta > 0$ . Suppose Assumptions 1, 3 and 4 hold, suppose (4.47) holds, and suppose that  $\alpha$  is chosen to satisfy

$$\alpha < \frac{1}{1 + \mathcal{TU}(2\rho)}. \quad (\text{B.116})$$

Then, in the proportional-growth asymptotic, ITP converges to an  $\alpha$ -stable point supported on some  $\Gamma_i$  such that  $i \in \Theta_2$ , with probability tending to 1 exponentially in  $n$ .

**Proof:** Given (B.116), we may apply Lemma 3.7 with  $\epsilon$  sufficiently small to deduce  $\alpha(1 + \mathcal{TU}_{2k}) < 1$ , with probability tending to 1 exponentially in  $n$ . Under Assumption 1, we may apply Lemma 4.3 and deduce convergence of ITP to an  $\alpha$ -stable point. We now show that this stable point must be supported on  $\Gamma_i$  such that  $i \in \Theta_2$ . For any  $\Gamma_i$  such that  $i \in \Theta_n^1$ , we have  $\Gamma_i \neq \Lambda$ , and we may therefore combine Theorem 4.2 with Lemma 4.4 to deduce that a necessary condition for there to be an  $\alpha$ -stable point on  $\Gamma_i$  is

$$\begin{aligned} & \|x_{\Lambda \setminus \Gamma_i}^*\| \cdot \sqrt{F_{\Gamma_i}} + \|x_{\Gamma_i \setminus \Lambda}^*\| + \sigma \cdot \sqrt{G_{\Gamma_i}} \\ & \geq \alpha \left[ \left( \frac{n-k}{n} \right) \|x_{\Lambda \setminus \Gamma_i}^*\| \cdot R_{\Gamma_i} - \sigma \sqrt{\frac{k(n-k)}{n^2}} (S_{\Gamma_i})(T_{\Gamma_i}) \right], \end{aligned} \quad (\text{B.117})$$

where

$$\begin{aligned} F_{\Gamma_i} & \sim \frac{k}{n-k+1} F(k, n-k+1); & G_{\Gamma_i} & \sim \frac{k}{n-k+1} F(k, n-k+1); \\ R_{\Gamma_i} & \sim \frac{1}{n-k} \chi_{n-k}^2; & S_{\Gamma_i} & \sim \frac{1}{n-k} \chi_{n-k}^2; & T_{\Gamma_i} & \sim \frac{1}{k} \chi_k^2. \end{aligned}$$

We also have, by (B.115),

$$\sigma \leq \frac{\|x_{\Lambda \setminus \Gamma_i}^*\|}{a^*(\rho; \zeta)} \quad (\text{B.118})$$

for any  $\Gamma_i$  such that  $i \in \Theta_n^1$ . Since  $\Gamma_i \neq \Lambda$ ,  $\|x_{\Lambda \setminus \Gamma_i}^*\| > 0$ , and substitution of (B.118) into (B.117), rearrangement and division by  $\|x_{\Lambda \setminus \Gamma_i}^*\|$  yields

$$a^*(\rho; \zeta) \left[ \alpha \left( \frac{n-k}{n} \right) \cdot R_{\Gamma_i} - \sqrt{F_{\Gamma_i}} \right] \leq \sqrt{G_{\Gamma_i}} + \alpha \sqrt{\frac{k(n-k)}{n^2}} \cdot S_{\Gamma_i} \cdot T_{\Gamma_i}.$$

Consequently,

$$\mathbb{P}(\overline{NSP}_\alpha) = \mathbb{P}\{\cup_{i \in \Theta_1} (\exists \text{ an } \alpha\text{-stable point supported on } \Gamma_i)\} \quad (\text{B.119})$$

$$= \mathbb{P}\left\{\bigcup_{i \in \Theta_1} \left[ a^*(\rho; \zeta) \left[ \alpha(1 - \rho_n) \cdot R_{\Gamma_i} - \sqrt{F_{\Gamma_i}} \right] \leq 1 + \sqrt{G_{\Gamma_i}} + \alpha \sqrt{\rho_n(1 - \rho_n)(S_{\Gamma_i})(T_{\Gamma_i})} \right] \right\}, \quad (\text{B.120})$$

where we write  $\rho_n$  for the sequence of values of the ratio  $k/n$ . For brevity's sake, let us define

$$\Phi[\rho, F, G, R, S, T] := 1 + \sqrt{G} + \alpha \sqrt{\rho(1 - \rho)(S)(T)} - a^*(\rho; \zeta) \cdot \left[ \alpha(1 - \rho) \cdot R - \sqrt{F} \right], \quad (\text{B.121})$$

so that (B.119) may equivalently be written as

$$\mathbb{P}(\overline{NSP}_\alpha) = \mathbb{P}\{\cup_{i \in \Theta_1} (\Phi[\rho_n, a^*(\rho; \zeta), F_{\Gamma_i}, G_{\Gamma_i}, R_{\Gamma_i}, S_{\Gamma_i}, T_{\Gamma_i}] \geq 0)\}. \quad (\text{B.122})$$

Given some  $\epsilon > 0$ , we now define

$$F^* = G^* := \mathcal{TIF}(\rho) + \epsilon; \quad R^* := 1 - \mathcal{TIL}(\rho, 1 - \rho) - \epsilon; \quad S^* := 1 + \mathcal{TIU}(\rho, 1 - \rho) + \epsilon; \quad T^* := 1 + \mathcal{TIU}(\rho, \rho) + \epsilon, \quad (\text{B.123})$$

Using (B.123), we deduce from (B.122) that

$$\begin{aligned} & \mathbb{P}(\overline{NSP}_\alpha) \\ & \leq \mathbb{P}\{\cup_{i \in \Theta_1} (\Phi[\rho_n, F_{\Gamma_i}, G_{\Gamma_i}, R_{\Gamma_i}, S_{\Gamma_i}, T_{\Gamma_i}] \geq \Phi[\rho_n, F^*, G^*, R^*, S^*, T^*])\} \end{aligned} \quad (\text{B.124})$$

$$+ \mathbb{P}\{\Phi[\rho_n, F^*, G^*, R^*, S^*, T^*] \geq \Phi[\rho, F^*, G^*, R^*, S^*, T^*] + \epsilon\} \quad (\text{B.125})$$

$$+ \mathbb{P}\{\Phi[\rho, F^*, G^*, R^*, S^*, T^*] + \epsilon \geq 0\}, \quad (\text{B.126})$$

since the event in the right-hand side of (B.122) lies in the union of the three events in (B.124), (B.125) and (B.126). Now (B.126) is a deterministic event, and  $a^*(\rho; \zeta)$  has been defined in such a way that, for any  $\zeta > 0$ , provided  $\epsilon$  is taken sufficiently small, the event has probability 0. This follows from (4.47), (4.48), B.114, and by the continuity of  $\Phi$ . The event (B.125) is also deterministic, and by continuity and since  $\rho_n \rightarrow \rho$ , it follows that there exists some  $\tilde{n}$  such that

$$\mathbb{P}\{\Phi[\rho_n, F^*, G^*, R^*, S^*, T^*] \geq \Phi[\rho, F^*, G^*, R^*, S^*, T^*] + \epsilon\} = 0 \quad \text{for all } n \geq \tilde{n}.$$

Taking limits as  $n \rightarrow \infty$ , the terms (B.125) and (B.126) are zero, leaving only (B.124), and we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(\overline{NSP}_\alpha) \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P}\{\cup_{i \in \Theta_1} (\Phi[\rho_n, a^*(\rho; \zeta), F_{\Gamma_i}, G_{\Gamma_i}, R_{\Gamma_i}, S_{\Gamma_i}, T_{\Gamma_i}] \geq \Phi[\rho_n, a^*(\rho; \zeta), F^*, G^*, R^*, S^*, T^*])\} \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P}\{\cup_{i \in \Theta_1} (F_{\Gamma_i} \geq F^*)\} + \lim_{n \rightarrow \infty} \mathbb{P}\{\cup_{i \in \Theta_1} (G_{\Gamma_i} \geq G^*)\} + \lim_{n \rightarrow \infty} \mathbb{P}\{\cup_{i \in \Theta_1} (R_{\Gamma_i} \leq R^*)\} \\ & + \lim_{n \rightarrow \infty} \mathbb{P}\{\cup_{i \in \Theta_1} (S_{\Gamma_i} \geq S^*)\} + \lim_{n \rightarrow \infty} \mathbb{P}\{\cup_{i \in \Theta_1} (T_{\Gamma_i} \geq T^*)\}, \end{aligned} \quad (\text{B.127})$$

where the last line follows from the monotonicity of  $\Phi$  with respect to  $F, G, R, S$  and  $T$ . Since  $\Theta_1 \subseteq \mathcal{T}_k$ , we may apply Lemmas 4.7 and 4.8 to (B.127), and since  $\Theta_1$  and  $\Theta_2$  partition  $\mathcal{T}_k$ , the result follows.  $\square$

Next we show that all  $\alpha$ -stable points supported on some  $\Gamma_i \in \Theta_2$  have bounded approximation error.

**Lemma B.7** *Suppose Assumptions 1, 3 and 4 hold, suppose that (4.47) holds, and suppose that  $\alpha$  is chosen to satisfy (B.116). There exists  $\zeta$  sufficiently small such that, in the proportional-growth asymptotic, any  $\alpha$ -stable point  $\bar{x}$  of ITP on  $\Gamma_i$  such that  $i \in \Theta_2$  satisfies*

$$\|\bar{x} - x^*\| \leq \xi(\rho) \cdot \sigma, \quad (\text{B.128})$$

where  $\xi(\rho)$  is defined in (4.49).

**Proof:** Suppose  $\bar{x}$  is a minimum-norm solution on  $\Gamma$ , so that  $\bar{x}_\Gamma = A_\Gamma^\dagger b$  and  $\bar{x}_{\Gamma^c} = 0$ . Then, using  $A_\Gamma^\dagger A_\Gamma = I$ , we have

$$\begin{aligned} (\bar{x} - x^*)_\Gamma &= A_\Gamma^\dagger (A_\Gamma x_\Gamma^* + A_{\Gamma^c} x_{\Gamma^c}^* + e) - x_\Gamma^* \\ &= x_\Gamma^* + A_\Gamma^\dagger (A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^* + A_{(\Lambda \cup \Gamma)^c} x_{(\Lambda \cup \Gamma)^c}^* + e) - x_\Gamma^* \\ &= A_\Gamma^\dagger (A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^* + e) + x_\Gamma^* - x_\Gamma^* \\ &= A_\Gamma^\dagger (A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^* + e), \end{aligned} \quad (\text{B.129})$$

while

$$(\bar{x} - x^*)_{\Gamma^c} = -x_{\Gamma^c}^*. \quad (\text{B.130})$$

Combining (B.129) and (B.130) using the triangle inequality, we may bound

$$\begin{aligned} \|\bar{x} - x^*\|^2 &\leq \|(\bar{x} - x^*)_{\Gamma}\|^2 + \|(\bar{x} - x^*)_{\Gamma^c}\|^2 \\ &= \|A_{\Gamma}^{\dagger}(A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^* + e)\|^2 + \|x_{\Gamma^c}^*\|^2 \\ &\leq \left[ \|A_{\Gamma}^{\dagger} A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^*\| + \|A_{\Gamma}^{\dagger} e\| \right]^2 + \|x_{\Lambda \setminus \Gamma}^*\|^2 + \|x_{(\Lambda \cup \Gamma)^c}^*\|^2. \end{aligned} \quad (\text{B.131})$$

We may deduce, by (4.36) of Lemma 4.4,

$$\|A_{\Gamma}^{\dagger} A_{\Lambda \setminus \Gamma} x_{\Lambda \setminus \Gamma}^*\|^2 = \|x_{\Lambda \setminus \Gamma}^*\|^2 \cdot P_{\Gamma}, \quad \text{where } P_{\Gamma} \sim \frac{k}{n-k+1} F(k, n-k+1), \quad (\text{B.132})$$

and by (4.38) of Lemma 4.4,

$$\|A_{\Gamma}^{\dagger} e\|^2 = \sigma^2 \cdot Q_{\Gamma}, \quad \text{where } Q_{\Gamma} \sim \frac{k}{n-k+1} F(k, n-k+1). \quad (\text{B.133})$$

Substituting (B.132) and (B.133) into (B.131), we have

$$\|\bar{x} - x^*\|^2 \leq \left[ \|x_{\Lambda \setminus \Gamma}^*\| \cdot \sqrt{P_{\Gamma}} + \sigma \cdot \sqrt{Q_{\Gamma}} \right]^2 + \|x_{\Lambda \setminus \Gamma}^*\|^2, \quad (\text{B.134})$$

and we may use (B.115) to further deduce

$$\begin{aligned} \|\bar{x} - x^*\|^2 &\leq \sigma^2 \left[ a^*(\rho; \zeta) \cdot \sqrt{P_{\Gamma}} + \sqrt{Q_{\Gamma}} \right]^2 + [a^*(\rho; \zeta)]^2 \cdot \sigma^2 \\ &= \sigma^2 \left\{ \left[ a^*(\rho; \zeta) \cdot \sqrt{P_{\Gamma}} + \sqrt{Q_{\Gamma}} \right]^2 + [a^*(\rho; \zeta)]^2 \right\}. \end{aligned} \quad (\text{B.135})$$

For the sake of brevity, let us define

$$\Psi(P, Q) := \sqrt{\left( a^*(\rho; \zeta) \cdot \sqrt{P} + \sqrt{Q} \right)^2 + [a^*(\rho; \zeta)]^2}, \quad (\text{B.136})$$

so that (B.135) may equivalently be written as

$$\|\bar{x} - x^*\| \leq \sigma \cdot \Psi[P_{\Gamma}, Q_{\Gamma}]. \quad (\text{B.137})$$

Given  $\zeta > 0$ , let us define

$$P^* = Q^* := \mathcal{TLF}(\rho) + \zeta. \quad (\text{B.138})$$

Now we use (B.137) to perform a union bound over all  $\Gamma_i$  such that  $i \in \Theta_2$ , writing  $\bar{x}_i$  for the minimum-norm solution supported on  $\Gamma_i$ , giving

$$\begin{aligned} &\mathbb{P} \{ \exists \text{ some } \Gamma_i \text{ such that } i \in \Theta_2 \text{ and } \|\bar{x}_i - x^*\| \geq \sigma \cdot \Psi[P^*, Q^*] \} \\ &= \mathbb{P} \left\{ \bigcup_{i \in \Theta_2} (\|\bar{x}_i - x^*\| \geq \sigma \cdot \Psi[P^*, Q^*]) \right\} \end{aligned} \quad (\text{B.139})$$

$$\leq \mathbb{P} \left\{ \bigcup_{i \in \Theta_2} (\|\bar{x}_i - x^*\| \geq \sigma \cdot \Psi[P_{\Gamma_i}, Q_{\Gamma_i}]) \right\} \quad (\text{B.140})$$

$$+ \mathbb{P} \left\{ \bigcup_{i \in \Theta_2} (\sigma \cdot \Psi[P_{\Gamma_i}, Q_{\Gamma_i}] \geq \sigma \cdot \Psi[P^*, Q^*]) \right\}, \quad (\text{B.141})$$

since the event in (B.139) lies in the union of the two events in (B.140) and (B.141). It is an immediate consequence of (B.135) that the event in (B.140) has probability 0. Taking limits of (B.141) as  $n \rightarrow \infty$ , we have

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P} \{ \exists \text{ some } \Gamma_i \text{ such that } i \in \Theta_2 \text{ and } \|\bar{x}_i - x^*\| \geq \sigma \cdot \Psi[P^*, Q^*] \} \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_2} (\sigma \cdot \Psi[P_{\Gamma_i}, Q_{\Gamma_i}] \geq \sigma \cdot \Psi[P^*, Q^*]) \right\} \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P} \{ \cup_{i \in \Theta_2} (P_{\Gamma_i} \geq P^*) \} + \lim_{n \rightarrow \infty} \mathbb{P} \{ \cup_{i \in \Theta_2} (Q_{\Gamma_i} \geq Q^*) \}, \end{aligned} \quad (\text{B.142})$$

where we used the monotonicity of  $\Psi$  with respect to  $P$  and  $Q$  in the last line. Since  $\Theta_2 \subseteq \mathcal{T}_k$ , and using (B.132) and (B.133), we may apply Lemma 4.8 to (B.142), yielding that each of the limits in the right-hand side of (B.142) converges to zero exponentially in  $n$ , and so finally

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \exists \text{ some } \Gamma_i \text{ such that } i \in \Theta_n^2 \text{ and } \|\bar{x}_i - x^*\| > \sigma \cdot \Psi[a^*(\rho; \zeta), P^*, Q^*] \right\} = 0,$$

exponentially in  $n$ . Since, by Lemma B.1, any stable point is necessarily a minimum-norm solution, and recalling the definition of  $a^*(\rho; \zeta)$  in (B.114),  $\Psi(P, Q)$  in (B.136), and the definitions of  $P^*$  and  $Q^*$  in (B.138), we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \exists \text{ some } \alpha\text{-stable point } \bar{x}_i \text{ on } \Gamma_i \text{ such that } i \in \Theta_n^2 \text{ and } \|\bar{x}_i - x^*\| \geq \sigma \cdot \Psi[P^*, Q^*] \right\} = 0,$$

with convergence exponential in  $n$ . Finally, by continuity,

$$\begin{aligned} \|\bar{x}_i - x\| &> \sigma \sqrt{\mathcal{T}\mathcal{L}\mathcal{F}(\rho)[1 + a(\rho)]^2 + [a(\rho)]^2} \\ \implies \|\bar{x}_i - x^*\| &\geq \sigma \sqrt{\mathcal{T}\mathcal{L}\mathcal{F}(\rho)[1 + a(\rho) + \zeta]^2 + [a(\rho) + \zeta]^2}, \end{aligned}$$

for some  $\zeta$  suitably small, and the result now follows from the definition of  $\xi_{SP}^{ITP\alpha}(\rho)$  in (4.49).  $\square$

It is now straightforward to prove the two main results for ITP.

**Proof of Theorem 4.11:** By Lemma B.6, we have convergence to an  $\alpha$ -stable point supported on some  $\Gamma_i$  such that  $i \in \Theta_2$ , to which we can apply Lemma B.7 deducing (B.128) with probability tending to 1 exponentially in  $n$ .  $\square$

**Proof of Corollary 4.12:** The result follows by setting  $\sigma := 0$  in Theorem 4.11.  $\square$

## B.2.2 Proof of NITP recovery results

In the case of NITP, it is possible to prove convergence to an  $\underline{\alpha}(\rho; \epsilon)$ -stable point, where

$$\underline{\alpha}(\rho; \epsilon) := \{\kappa[1 + \mathcal{T}\mathcal{U}(2\rho) + \epsilon]\}^{-1}, \quad (\text{B.143})$$

for some  $\epsilon > 0$ .

The proof of Theorem 4.15 for NITP takes broadly the same approach as for the corresponding result for ITP in Section B.2.1. However, in order to finally eliminate the dependence upon  $\epsilon$  in  $\underline{\alpha}(\rho; \epsilon)$ , the results corresponding to Lemmas B.6 and B.3 for ITP need to be combined together. This is accomplished by Lemma B.9, which establishes that, provided (4.53) holds and  $\epsilon$  is taken sufficiently small, NITP converges to an  $\underline{\alpha}(\rho; \epsilon)$ -stable point on some  $\Gamma_i$  such that  $i \in \Theta_n^2$  (the NITP support set partition is given in (B.145) below). Lemma B.10 corresponds to Lemma B.7 for ITP, giving bounds on the approximation error of an  $\underline{\alpha}(\rho; \epsilon)$ -stable point on some  $\Gamma_i$  such that  $i \in \Theta_n^2$ , for any  $\epsilon > 0$ . Combining the two lemmas leads us to conclude that NITP converges to some limit point with bounded approximation error. We write  $NSP_{\underline{\alpha}}$  for the event that there is no  $\underline{\alpha}(\rho; \epsilon)$ -stable point on any  $\Gamma_i$  such that  $i \in \Theta_n^1$ .

We next introduce the support set partition definition relevant for NITP.

**Definition B.8 (Support set partition for NITP)** *Suppose  $\rho \in (0, 1/2]$ . Given  $\zeta > 0$ , let us write*

$$a^*(\rho; \zeta) := a(\rho) + \zeta, \quad (\text{B.144})$$

where  $a(\rho)$  is defined in (4.54), let us write  $\{\Gamma_i : i \in S_n\}$  for the set of all possible support sets of cardinality  $k$ , and let us disjointly partition  $S_n := \Theta_n^1 \cup \Theta_n^2$  such that

$$\Theta_n^1 := \left\{ i \in S_n : \|x_{\Lambda \setminus \Gamma_i}^*\| > \Sigma \cdot a^*(\rho; \zeta) \right\}; \quad \Theta_n^2 := \left\{ i \in S_n : \|x_{\Lambda \setminus \Gamma_i}^*\| \leq \Sigma \cdot a^*(\rho; \zeta) \right\}. \quad (\text{B.145})$$

**Lemma B.9** *Choose  $\zeta > 0$ . Suppose Assumptions 1, 3 and 4 hold, and suppose that (4.53) holds. Then there exists  $\epsilon$  such that, in the proportional-growth asymptotic, NITP converges to an  $\underline{\alpha}(\rho; \epsilon)$ -stable point on some  $\Gamma_i$  such that  $i \in \Theta_n^2$ , with probability tending to 1 exponentially in  $n$ .*

**Proof:** Under Assumption 1, we have by Theorem B.4 convergence of NITP to a  $[\kappa(1 + TU_{2k})]^{-1}$ -stable point. By Definition 4.1, for any  $\alpha_1 < \alpha_2$ , the set of  $\alpha_1$ -stable points includes the set of  $\alpha_2$ -stable points, and this observation combines with Lemma 3.7 to imply convergence to an  $\underline{\alpha}(\rho; \epsilon)$ -stable point, where  $\underline{\alpha}(\rho; \epsilon)$  is defined in (B.143), with probability tending to 1 exponentially in  $n$ . We now rehearse the argument of Lemma B.6 to show that, provided  $\epsilon$  is taken sufficiently small, this stable point must be on  $\Gamma_i$  such that  $i \in \Theta_n^2$ . For any  $\Gamma_i$  such that  $i \in \Theta_n^1$ , we have  $\Gamma_i \neq \Lambda$ , and we may therefore use Theorem 4.2 and Lemma 4.4 with  $\Gamma := \Gamma_i$  to deduce that, given some  $\epsilon > 0$ , a necessary condition for there to be an  $\underline{\alpha}(\rho; \epsilon)$ -stable point on  $\Gamma_i$  is

$$\begin{aligned} & \|x_{\Lambda \setminus \Gamma_i}^*\| \cdot \sqrt{F_{\Gamma_i}} + \sigma \cdot \sqrt{G_{\Gamma_i}} \\ & \geq \underline{\alpha}(\rho; \epsilon) \left[ \left( \frac{n-k}{n} \right) \|x_{\Lambda \setminus \Gamma_i}^*\| \cdot R_{\Gamma_i} - \sigma \cdot \sqrt{\frac{k(n-k)}{n^2} \cdot S_{\Gamma_i} \cdot T_{\Gamma_i}} \right], \end{aligned} \quad (\text{B.146})$$

where

$$\begin{aligned} F_{\Gamma_i} & \sim \frac{k}{n-k+1} \mathcal{F}(k, n-k+1); & G_{\Gamma_i} & \sim \frac{k}{n-k+1} \mathcal{F}(k, n-k+1); \\ R_{\Gamma_i} & \sim \frac{1}{n-k} \chi_{n-k}^2; & S_{\Gamma_i} & \sim \frac{1}{n-k} \chi_{n-k}^2; & T_{\Gamma_i} & \sim \frac{1}{k} \chi_k^2. \end{aligned}$$

We also have, by (B.145),

$$\sigma \leq \frac{\|x_{\Lambda \setminus \Gamma_i}^*\|}{a^*(\rho; \zeta)} \quad (\text{B.147})$$

for any  $\Gamma_i$  such that  $i \in \Theta_n^1$ . Since  $\Gamma_i \neq \Lambda$ ,  $\|x_{\Lambda \setminus \Gamma_i}^*\| > 0$ , and substitution of (B.147) into (B.146), rearrangement and division by  $\|x_{\Lambda \setminus \Gamma_i}^*\|$  yields

$$a^*(\rho; \zeta) \left[ \underline{\alpha}(\rho; \epsilon) \left( \frac{n-k}{n} \right) \cdot R_{\Gamma_i} - \sqrt{F_{\Gamma_i}} \right] \leq \sqrt{G_{\Gamma_i}} + \underline{\alpha}(\rho; \epsilon) \sqrt{\frac{k(n-k)}{n^2} \cdot S_{\Gamma_i} \cdot T_{\Gamma_i}},$$

and consequently

$$\begin{aligned} \mathbb{P}(\overline{NSP_{\underline{\alpha}(\delta, \rho; \zeta)}}) & = \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (\exists \text{ an } \underline{\alpha}(\rho; \epsilon)\text{-stable point supported on } \Gamma_i) \right\} \\ & \leq \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (\Phi[\rho_n, F_{\Gamma_i}, G_{\Gamma_i}, R_{\Gamma_i}, S_{\Gamma_i}, T_{\Gamma_i}] \geq 0) \right\}, \end{aligned} \quad (\text{B.148})$$

where we write  $\rho_n$  for the sequence of values of the ratio  $k/n$ , and where

$$\Phi[\rho, F, G, R, S, T] := \sqrt{G} + \underline{\alpha}(\rho; \epsilon) \sqrt{\rho(1-\rho)(S)(T)} - a^*(\rho; \zeta) \cdot \left[ \underline{\alpha}(\rho; \epsilon)(1-\rho) \cdot R - \sqrt{F} \right]. \quad (\text{B.149})$$

We now define

$$\begin{aligned} F^* & = G^* := \mathcal{TIF}(\rho) + \epsilon; & R^* & := 1 - \mathcal{TIL}(\rho, 1-\rho) - \epsilon; \\ S^* & := 1 + \mathcal{TIU}(\rho, 1-\rho) + \epsilon; & T^* & := 1 + \mathcal{TIU}(\rho, \rho) + \epsilon. \end{aligned} \quad (\text{B.150})$$

Using (B.150), we deduce from (B.148) that

$$\begin{aligned} & \mathbb{P}(\overline{NSP_{\underline{\alpha}(\delta, \rho; \zeta)}}) \\ & \leq \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (\Phi[\rho_n, F_{\Gamma_i}, G_{\Gamma_i}, R_{\Gamma_i}, S_{\Gamma_i}, T_{\Gamma_i}] \geq \Phi[\rho_n, F^*, G^*, R^*, S^*, T^*]) \right\} \end{aligned} \quad (\text{B.151})$$

$$+ \mathbb{P} \left\{ \Phi[\rho_n, F^*, G^*, R^*, S^*, T^*] \geq \Phi[\rho, F^*, G^*, R^*, S^*, T^*] + \epsilon \right\} \quad (\text{B.152})$$

$$+ \mathbb{P} \left\{ \Phi[\rho, F^*, G^*, R^*, S^*, T^*] + \epsilon \geq 0 \right\}, \quad (\text{B.153})$$

since the event in (B.148) lies in the union of the three events in (B.151), (B.152) and (B.153). Now (B.153) is a deterministic event, and  $a^*(\rho; \zeta)$  has been defined in such a way that, for any  $\zeta > 0$ , provided  $\epsilon$  is taken sufficiently small, the event has probability 0. This follows from (4.53), (4.54), (B.144), and by the continuity of  $\Phi$ . The event (B.152) is also deterministic, and by continuity and since  $\rho_n \rightarrow \rho$ , it follows that there exists some  $\tilde{n}$  such that

$$\mathbb{P} \left\{ \Phi[\rho_n, F^*, G^*, R^*, S^*, T^*] \geq \Phi[\rho, F^*, G^*, R^*, S^*, T^*] + \epsilon \right\} = 0 \quad \text{for all } n \geq \tilde{n}.$$

Taking limits as  $n \rightarrow \infty$ , the terms (B.152) and (B.153) are zero, leaving only (B.151), and we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(\overline{NSP_{\underline{\alpha}(\delta, \rho; \zeta)}}) \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (\Phi[\rho_n, F_{\Gamma_i}, G_{\Gamma_i}, R_{\Gamma_i}, S_{\Gamma_i}, T_{\Gamma_i}] \geq \Phi[\rho_n, F^*, G^*, R^*, S^*, T^*]) \right\} \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (F_{\Gamma_i} \geq F^*) \right\} + \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (G_{\Gamma_i} \geq G^*) \right\} + \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (R_{\Gamma_i} \leq R^*) \right\} \\ & + \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (S_{\Gamma_i} \geq S^*) \right\} + \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^1} (T_{\Gamma_i} \geq T^*) \right\}, \end{aligned} \quad (\text{B.154})$$

where the last line follows from the monotonicity of  $\Phi$  with respect to  $F, G, R, S$  and  $T$ . Since  $\Theta_n^1 \subseteq S_n$ , we may apply Lemmas 4.7 and 4.8 to (B.154), and we deduce  $\mathbb{P}(\overline{NSP}_{\underline{\alpha}(\delta, \rho; \zeta)}) \rightarrow 0$  as  $n \rightarrow \infty$ , exponentially in  $n$ , as required.  $\square$

**Lemma B.10** *Suppose Assumptions 1, 3 and 4 hold, and suppose that (4.53) holds. Given any  $\epsilon > 0$ , there exists  $\zeta$  sufficiently small such that, in the proportional-growth asymptotic, any  $\underline{\alpha}(\rho; \epsilon)$ -stable point on  $\Gamma_i$  such that  $i \in \Theta_n^2$  satisfies (4.56), with probability tending to 1 exponentially in  $n$ .*

**Proof:** Suppose  $\bar{x}$  is a minimum-norm solution on  $\Gamma$ , so that  $\bar{x}_\Gamma = A_\Gamma^\dagger b$ . Then we may follow the argument of Lemma B.7 to deduce (B.134), where

$$P_\Gamma \sim \frac{k}{n-k+1} \mathcal{F}(k, n-k+1); \quad Q_\Gamma \sim \frac{k}{n-k+1} \mathcal{F}(k, n-k+1). \quad (\text{B.155})$$

Combining (B.134) with (B.145), we may further deduce

$$\begin{aligned} \|\bar{x} - x^*\|^2 &\leq \sigma^2 \left[ a^*(\rho; \zeta) \cdot \sqrt{P_\Gamma} + \sqrt{Q_\Gamma} \right]^2 + [a^*(\rho; \zeta)]^2 \cdot \sigma^2 \\ &= \sigma^2 \left\{ \left[ a^*(\rho; \zeta) \cdot \sqrt{P_\Gamma} + \sqrt{Q_\Gamma} \right]^2 + [a^*(\rho; \zeta)]^2 \right\}. \end{aligned} \quad (\text{B.156})$$

For the sake of brevity, let us define

$$\Psi[P, Q] := \sqrt{\left( a^*(\rho; \zeta) \cdot \sqrt{P} + \sqrt{Q} \right)^2 + a^*(\rho; \zeta)^2}, \quad (\text{B.157})$$

so that (B.156) may equivalently be written as

$$\|\bar{x} - x^*\| \leq \sigma \cdot \Psi[P_\Gamma, Q_\Gamma]. \quad (\text{B.158})$$

First suppose that  $\sigma > 0$ . Given  $\zeta > 0$ , let us define

$$P^* = Q^* := \mathcal{TLF}(\rho) + \zeta. \quad (\text{B.159})$$

Now we use (B.158) to perform a union bound over all  $\Gamma_i$  such that  $i \in \Theta_n^2$ , writing  $\bar{x}_i$  for the minimum-norm solution on  $\Gamma_i$ , giving

$$\begin{aligned} &\mathbb{P} \left\{ \exists \text{ some } \Gamma_i \text{ such that } i \in \Theta_n^2 \text{ and } \|\bar{x}_i - x^*\| > \sigma \cdot \Psi[P^*, Q^*] \right\} \\ &= \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^2} (\|\bar{x}_i - x^*\| > \sigma \cdot \Psi[P^*, Q^*]) \right\} \end{aligned} \quad (\text{B.160})$$

$$\leq \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^2} (\|\bar{x}_i - x^*\| > \sigma \cdot \Psi[P_{\Gamma_i}, Q_{\Gamma_i}]) \right\} \quad (\text{B.161})$$

$$+ \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^2} (\sigma \cdot \Psi[P_{\Gamma_i}, Q_{\Gamma_i}] \geq \sigma \cdot \Psi[P^*, Q^*]) \right\}, \quad (\text{B.162})$$

since the event in (B.160) lies in the union of the two events in (B.161) and (B.162). It is an immediate consequence of (B.158) that the event in (B.161) has probability 0. Taking limits of (B.162) as  $n \rightarrow \infty$ , and cancelling  $\sigma$ , we have

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \exists \text{ some } \Gamma_i \text{ such that } i \in \Theta_n^2 \text{ and } \|\bar{x}_i - x^*\| > \sigma \cdot \Psi[P^*, Q^*] \right\} \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^2} (\Psi[P_{\Gamma_i}, Q_{\Gamma_i}] \geq \Psi[P^*, Q^*]) \right\} \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^2} (P_{\Gamma_i} \geq P^*) \right\} + \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i \in \Theta_n^2} (Q_{\Gamma_i} \geq Q^*) \right\}, \end{aligned} \quad (\text{B.163})$$

where we used the monotonicity of  $\Psi$  with respect to  $P$  and  $Q$  in the last line. Since  $\Theta_n^2 \subseteq S_n$ , and using (B.155), we may apply Lemma 4.8 to (B.163), yielding that each of the limits in the right-hand side of (B.163) converges to zero exponentially in  $n$ , and so finally

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \exists \text{ some } \Gamma_i \text{ such that } i \in \Theta_n^2 \text{ and } \|\bar{x}_i - x^*\| > \sigma \cdot \Psi[P^*, Q^*] \right\} = 0,$$

with convergence at a rate exponential in  $n$  also by Lemma 4.8. The same result also holds when  $\sigma = 0$  by (B.156). Since by Lemma B.1, any stable point is necessarily a minimum-norm solution, and recalling the definition of  $\Psi(P, Q)$  in (B.157), and the definitions of  $P^*$ ,  $Q^*$  in (B.159), we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \begin{array}{l} \exists \text{ some } \underline{\alpha}\text{-stable point } \bar{x}_i \text{ on } \Gamma_i \text{ such that } i \in \Theta_n^2 \text{ and} \\ \|\bar{x}_i - x^*\| > \sigma \sqrt{\mathcal{TIF}(\rho) [1 + a(\rho) + \zeta]^2 + [a(\rho) + \zeta]^2} \end{array} \right\} = 0, \quad (\text{B.164})$$

with convergence exponential in  $n$ . Finally, by continuity,

$$\begin{aligned} \|\bar{x}_i - x^*\| &> \sigma \sqrt{\mathcal{TIF}(\rho) [1 + a(\rho)]^2 + 1 + [a(\rho)]^2} \\ \implies \|\bar{x}_i - x^*\| &> \sigma \sqrt{\mathcal{TIF}(\rho) [1 + a(\rho) + \zeta]^2 + [a(\rho) + \zeta]^2}, \end{aligned}$$

for some  $\zeta$  suitably small, and the result now follows from the definition of  $\xi(\rho)$  in (4.55).  $\square$

It is now straightforward to prove the two main results for NITP.

**Proof of Theorem 4.15:** By Lemma B.9, there exists  $\epsilon > 0$  such that N-IHT converges to an  $\underline{\alpha}(\delta, \rho; \epsilon)$ -stable point on some  $\Gamma_i$  such that  $i \in \Theta_n^2$ , and for this choice of  $\epsilon$ , we can apply Lemma B.10 to deduce the result.  $\square$

**Proof of Corollary 4.16:** The result follows by setting  $\sigma := 0$  in Theorem 4.15.  $\square$

## C Large deviations results in the tree-based setting

This appendix develops large deviation bounds in the simplified proportional-growth asymptotic of Definition 1.1 for various quantities related to Gaussian matrices, which are required to hold for all permissible support sets.

In what follows, let the tree order  $d$  to be some fixed integer with  $d \geq 2$ . We need to count  $|\mathcal{T}_k|$ , the number of permissible support sets in the  $d$ -ary tree-based framework, which is bounded above by  $T(k)$ , the total number of ordered, rooted  $d$ -ary trees of cardinality  $k$ . Recalling Lemma 3.4, we have

$$T(k) = \frac{1}{(d-1)k+1} \binom{dk}{k}.$$

A similar result was proved in [1, Proposition 1] for the case of binary trees ( $d = 2$ ), though the result given above represents a generalization to any  $d > 2$ , and in fact also gives a tightening of the result in [1] in the case where  $\log_2(N) > k$ . Note also that we have an upper bound on  $|\mathcal{T}_k|$  which is independent of  $N$ . This is in contrast to the total number of supports, which is  $\binom{N}{k}$ . However,  $|\mathcal{T}_k|$  may not attain this upper bound if additional structure is imposed. In a typical wavelet tree model, for example, the root node has only  $d-1$  children [13]). In addition, the number of levels in a wavelet tree structure is typically limited to  $J = \log_d(N)$ , which represents a further restriction if  $\log_d(N) < k$ . It follows that, while it is possible to give an upper bound on  $|\mathcal{T}_k|$  which is valid for any  $N$ ,  $|\mathcal{T}_k|$  does in general depend on both  $k$  and  $N$ .

We will make use of the following limiting result for  $T(k)$ .

**Lemma C.1 (Tree counting limit)**

$$\lim_{k \rightarrow \infty} \frac{1}{k} \ln T(k) = d \cdot H(d^{-1}), \quad (\text{C.165})$$

where  $H(\cdot)$  is defined in (3.13).

**Proof:**

$$\begin{aligned}
\lim_{k \rightarrow \infty} \frac{1}{k} \ln T(k) &= \lim_{k \rightarrow \infty} \frac{1}{k} \ln \left[ \frac{1}{(d-1)k+1} \binom{dk}{k} \right] \\
&= \lim_{k \rightarrow \infty} \frac{1}{k} \ln \left[ \frac{1}{(d-1)k+1} \right] + \lim_{k \rightarrow \infty} \frac{1}{k} \ln \binom{dk}{k} \\
&= 0 + \lim_{k \rightarrow \infty} d \cdot \frac{1}{dk} \ln \binom{dk}{k} \\
&= d \cdot H(d^{-1}),
\end{aligned}$$

where the last step follows from Stirling's formula.  $\square$

We proceed to proving the validity of the bounds on tree-based RIP constants for Gaussian matrices given in Definition 3.1.

**Proof of Lemma 3.7:** We follow closely the method of proof in [3, Proposition 2.6]. That there is a unique solution to (3.16) follows entirely as in the proof of [3, Proposition 2.6]. Select  $\epsilon > 0$  and let  $(k, n)$  be such that  $k/n = \rho_n$ . Then

$$\begin{aligned}
\mathbb{P}[TU_k \geq \mathcal{TU}(\rho_n) + \epsilon] &= \mathbb{P}[TU_k \geq \lambda^{max}(\rho_n) - 1 + \epsilon] \\
&= \mathbb{P}[1 + TU_k \geq \lambda^{max}(\rho_n) + \epsilon]
\end{aligned} \tag{C.166}$$

Equivalent to (C.166) is the requirement that the maximum eigenvalue of at least one of the Wishart matrices  $A_\Gamma^T A_\Gamma$  is greater than or equal to  $\lambda^{max}(\rho_n) + \epsilon$ , considering all  $\Gamma \in \mathcal{T}_k$ . Writing  $f_{max}(k, n; \lambda)$  for the pdf of a  $k \times k$  Wishart matrix distributed as  $\mathcal{W}_k(n, 1/n)$ , we may therefore perform a union bound over all  $\Gamma \in \mathcal{T}_k$  and write

$$\mathbb{P}[TU_k \geq \mathcal{TU}(\rho_n) + \epsilon] \leq |\mathcal{T}_k| \int_{\lambda^{max}(\rho_n) + \epsilon}^{\infty} f_{max}(k, n; \lambda) d\lambda. \tag{C.167}$$

Now, defining

$$g_{max}(k, n; \lambda) = \left[ (2\pi)^{1/2} (n\lambda)^{-3/2} \left( \frac{n\lambda}{2} \right)^{(n+k)/2} \frac{1}{\Gamma(\frac{k}{2}) \Gamma(\frac{n}{2})} \right] \cdot e^{-n\lambda/2},$$

we may use [3, Lemma 2.4] to deduce

$$\begin{aligned}
f_{max}(k, n; \lambda) &\leq g_{max}(k, n; \lambda) \\
&= \phi(n, \rho_n) \lambda^{-\frac{3}{2}} \lambda^{\frac{n}{2}(1+\rho_n)} e^{-\frac{n}{2}\lambda}
\end{aligned} \tag{C.168}$$

where

$$\phi(n, \rho_n) = (2\pi)^{1/2} n^{-3/2} \frac{n^{\frac{n}{2}(1+\rho_n)}}{2} \frac{1}{\Gamma(\frac{n}{2}\rho_n) \Gamma(\frac{n}{2})}.$$

Since  $\lambda^{max}(\rho_n) > 1 + \rho_n$ , the quantity  $\lambda^{\frac{n}{2}(1+\rho_n)} e^{-\frac{n}{2}\lambda}$  is strictly decreasing in  $\lambda$  on  $[\lambda^{max}(\rho_n), \infty)$ . Combining with (C.168), we therefore have

$$\begin{aligned}
\int_{\lambda^{max}(\rho_n) + \epsilon}^{\infty} f_{max}(k, n; \lambda) d\lambda &\leq \phi(n, \rho_n) [\lambda^{max}(\rho_n) + \epsilon]^{\frac{n}{2}(1+\rho_n)} e^{-\frac{n}{2}[\lambda^{max}(\rho_n) + \epsilon]} \int_{\lambda^{max}(\rho_n) + \epsilon}^{\infty} \lambda^{-\frac{3}{2}} d\lambda \\
&= [\lambda^{max}(\rho_n) + \epsilon]^{\frac{3}{2}} g_{max}[k, n; \lambda^{max}(\rho_n) + \epsilon] \int_{\lambda^{max}(\rho_n) + \epsilon}^{\infty} \lambda^{-\frac{3}{2}} d\lambda \\
&= 2[\lambda^{max}(\rho_n) + \epsilon] g_{max}[k, n; \lambda^{max}(\rho_n) + \epsilon],
\end{aligned}$$

which may be substituted into (C.167) to obtain

$$\mathbb{P}[TU_k \geq \mathcal{TU}(\rho_n) + \epsilon] \leq 2|\mathcal{T}_k| [\lambda^{max}(\rho_n) + \epsilon] g_{max}[k, n; \lambda^{max}(\rho_n) + \epsilon],$$

and we furthermore apply [3, Lemma 2.5] to give

$$\mathbb{P}[TU_k \geq \mathcal{TU}(\rho_n) + \epsilon] \leq 2|\mathcal{T}_k| [\lambda^{max}(\rho_n) + \epsilon] p_{max}[n, \lambda^{max}(\rho_n) + \epsilon] \exp\{n \cdot \psi_{max}[\lambda^{max}(\rho_n) + \epsilon, \rho]\}, \tag{C.169}$$

where  $p_{max}(n, \lambda)$  is a polynomial in  $n$  and  $\lambda$ . Now we may take limits of both sides of (C.169), using (C.165), to deduce

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}[TU_k \geq \mathcal{TU}(\rho_n) + \epsilon] \leq d\rho \cdot H(d^{-1}) + \psi_{max}[\lambda^{max}(\rho) + \epsilon, \rho],$$



which is strictly negative by Definition 3.6, from which (3.18) now follows using the same argument as in the proof of lemma 4.7. An analogous argument may be followed to prove (3.19) which we omit for the sake of brevity.  $\square$

The large deviations bounds for the  $\chi^2$  and  $\mathcal{F}$  distributions follow.

**Proof of Lemma 4.7:** By [14, Lemma A.2], we have for all  $i \in S_n$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}(X_l^i \geq 1 + \nu) \leq -\frac{\gamma}{2}[\nu - \ln(1 + \nu)]. \quad (\text{C.170})$$

Union bounding  $\mathbb{P}(X_l^i \geq 1 + \nu)$  over all  $i \in S_n$  gives

$$\mathbb{P}\{\cup_{i \in S_n} (X_l^i \geq 1 + \nu)\} \leq \sum_i \mathbb{P}(X_l^i \geq 1 + \nu) \leq T(k) \cdot \mathbb{P}(X_l^1 \geq 1 + \nu). \quad (\text{C.171})$$

Taking logarithms and limits of the right-hand side of (C.171), using (C.170) and (C.165), we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln [T(k) \cdot \mathbb{P}(X_l^1 \geq 1 + \nu)] = d\rho \cdot H(d^{-1}) - \frac{\lambda}{2}[\nu - \ln(1 + \nu)],$$

and so (C.171) implies that, for any  $\eta > 0$ ,

$$\frac{1}{n} \ln \mathbb{P}\{\cup_{i \in S_n} (X_l^i \geq 1 + \nu)\} \leq d\rho \cdot H(d^{-1}) - \frac{\lambda}{2}[\nu - \ln(1 + \nu)] + \eta, \quad (\text{C.172})$$

for all  $n$  sufficiently large. By the definition of  $\mathcal{TU}(\rho, \lambda)$  in (4.40), and since  $[\nu - \ln(1 + \nu)]$  is strictly increasing on  $\nu > 0$ , then, for any  $\epsilon > 0$ , setting  $\nu := \nu^* = \mathcal{TU}(\rho, \lambda) + \epsilon$  and choosing  $\eta$  sufficiently small in (C.172) ensures

$$\frac{1}{n} \ln \mathbb{P}\{\cup_{i \in S_n} (X_l^i \geq 1 + \nu^*)\} \leq -c_Q \quad \text{for all } n \text{ sufficiently large,}$$

where  $c_Q$  is some positive constant, from which it follows that

$$\mathbb{P}\{\cup_{i \in S_n} (X_l^i \geq 1 + \nu^*)\} \leq e^{-c_Q \cdot n} \quad \text{for all } n \text{ sufficiently large,}$$

and (4.43) follows. Combining the same union bound argument with the lower tail result of [14, Lemma A.2] shows that, if we take  $\nu^* = \mathcal{TU}(\rho, \lambda) + \epsilon$  for some  $\epsilon > 0$ , then

$$\frac{1}{n} \ln \mathbb{P}\{\cup_{i \in S_n} (X_l^i \leq 1 - \nu^*)\} \leq -c_P \quad \text{for all } n \text{ sufficiently large,}$$

where  $c_P$  is some positive constant, and (4.44) follows similarly to (4.43).  $\square$

**Proof of Lemma 4.8:** By [14, Lemma A.5], we have for all  $i \in S_n$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}(X_n^i \geq f) \leq -\frac{1}{2}[\ln(1 + f) - \rho \ln f - H(\rho)]. \quad (\text{C.173})$$

Union bounding  $\mathbb{P}(X_n^i \geq f)$  over all  $i \in S_n$  gives

$$\mathbb{P}\{\cup_{i \in S_n} (X_n^i \geq f)\} \leq \sum_{i \in S_n} \mathbb{P}(X_n^i \geq f) = |S_n| \cdot \mathbb{P}(X_n^1 \geq f), \quad (\text{C.174})$$

Taking logarithms and limits of the right-hand side of (C.174), using (C.173) and (C.165), we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln [|S_n| \cdot \mathbb{P}(X_n^1 \geq f)] = d\rho \cdot H(d^{-1}) - \frac{1}{2}[\ln(1 + f) - \rho \ln f - H(\rho)],$$

which combines with (C.174) to imply that, for any  $\eta > 0$ ,

$$\frac{1}{n} \ln \mathbb{P}\{\cup_{i \in S_n} (X_n^i \geq f)\} \leq d\rho \cdot H(d^{-1}) - \frac{1}{2}[\ln(1 + f) - \rho \ln f - H(\rho)] + \eta, \quad (\text{C.175})$$

for all  $n$  sufficiently large. By the definition of  $\mathcal{TIF}(\rho)$  in (4.42), and since the left-hand side of (4.42) on  $f > \frac{\rho}{1 - \rho}$  is strictly increasing in  $f$ , then, for any  $\epsilon > 0$ , setting  $f := f^* = \mathcal{TIF}(\rho) + \epsilon$  and choosing  $\eta$  sufficiently small in (C.175) ensures

$$\frac{1}{n} \ln \mathbb{P}\{\cup_{i \in S_n} (X_n^i \geq f^*)\} \leq -c_I \quad \text{for all } n \text{ sufficiently large,}$$

where  $c_I$  is some positive constant, from which the result follows using the same argument as in the proof of lemma 4.7.  $\square$

## References

- [1] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56:1982–2001, 2010.
- [2] R. Baraniuk and D. Jones. A signal-dependent time-frequency representation: Fast algorithm for optimal kernel design. *IEEE Transactions on Signal Processing*, 42(12):3530–3535, 1994.
- [3] J. Blanchard, C. Cartis, and J. Tanner. Compressed sensing: How sharp is the restricted isometry property? *SIAM Review*, 53(1):105–125, 2011.
- [4] J. Blanchard, C. Cartis, J. Tanner, and A. Thompson. Phase transitions for greedy sparse approximation algorithms; extended technical report. Technical Report ERGO 09-010, School of Mathematics, University of Edinburgh, 2009.
- [5] J. Blanchard, C. Cartis, J. Tanner, and A. Thompson. Phase transitions for greedy sparse approximation algorithms. *Applied and Computational Harmonic Analysis*, 30(2):188–203, 2011.
- [6] J. Blanchard and A. Thompson. On support sizes of restricted isometry constants. *Applied and Computational Harmonic Analysis*, 29(3):382–390, 2010.
- [7] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and its Applications*, 14(5):629–654, 2008.
- [8] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.
- [9] T. Blumensath and M. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Transactions on Information Theory*, 55(4):1872–1882, 2009.
- [10] T. Blumensath and M. Davies. Normalised iterative hard thresholding: guaranteed stability and performance. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):298–309, 2010.
- [11] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.
- [12] E. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.
- [13] C. Cartis and A. Thompson. An exact tree projection algorithm for wavelets. *IEEE Signal Processing Letters*, 20(11):1028–1031, 2013.
- [14] C. Cartis and A. Thompson. A new and improved quantitative recovery analysis for iterative hard thresholding algorithms in compressed sensing. *IEEE Transactions on Information Theory*, 61(4):1–24, 2013.
- [15] D. Donoho. CART and best ortho-basis: A connection. *Annals of Statistics*, 25(5):1870–1911, 1997.
- [16] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [17] D. Donoho. High-dimensional centrosymmetric polytopes with neighborliness proportional to dimension. *Discrete and Computational Geometry*, 35(4):617–652, 2006.
- [18] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563, 2011.
- [19] R. Graham, D. Knuth, and O. Patashnik. *Concrete mathematics: a foundation for computer science*. Addison-Wesley, 1994.
- [20] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, third edition, 2009.
- [21] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2008.
- [22] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 1999.
- [23] A. Thompson. *Quantitative analysis of algorithms for compressed signal recovery*. PhD thesis, School of Mathematics, University of Edinburgh, 2012.