Daniel A. Spielman · Shang-Hua Teng

Smoothed Analysis of Termination of Linear Programming Algorithms

Abstract. We perform a smoothed analysis of a termination phase for linear programming algorithms. By combining this analysis with the smoothed analysis of Renegar's condition number by Dunagan, Spielman and Teng (`http://arxiv.org/abs/cs.DS/0302011`) we show that the smoothed complexity of interior-point algorithms for linear programming is $O(m^3 \log(m/\sigma))$. In contrast, the best known bound on the worst-case complexity of linear programming is $O(m^3 L)$, where L could be as large as m. We include an introduction to smoothed analysis and a tutorial on proof techniques that have been useful in smoothed analyses.

1. Introduction

This paper has two objectives: to provide an introduction to smoothed analysis and to present a new result—the smoothed analysis of a termination phase for linear programming algorithms. The termination phase is a simple heuristic for rounding a good solution of a linear program to an optimal solution of the program. Assuming the data $(A, \boldsymbol{b}, \boldsymbol{c})$ of a linear program is subject to perturbation, we analyze how good the solution must be for the rounding to succeed.

We begin with an intuitive introduction to smoothed analysis (Section 1.1) followed by a more formal introduction (Section 1.2). After introducing necessary notation in Section 2, we survey the complexity of interior-point algorithms (Section 3), emphasizing the role of condition numbers (Section 4). We then explain the termination algorithm (Section 5), present its smoothed analysis at a high level (Section 6), and then delve into the geometric (Section 7) and probabilistic (Section 8) aspects of its analysis. In Section 8, we include a tutorial on the fundamental techniques used in this work and in the smoothed analysis of the

Daniel A. Spielman: Department of Mathematics, Massachusetts Institute of Technology e-mail: `spielman@math.mit.edu`

Shang-Hua Teng: Department of Computer Science, Boston University and, Akamai Technologies Inc. e-mail: `steng@cs.bu.edu`

simplex method [ST01]. Finally, in Section 9, we explain how the analysis of termination is related to the analysis of the simplex method. We conclude with some open questions.

1.1. Intuitive Introduction to Smoothed Analysis

Folklore holds that most algorithms have much better performance in practice than can be proved theoretically. This is partially due to the lack of a theoretical definition of "practice", partially due to the approximations made in most theoretical analyses, and partially due to the dearth of performance measures considered in theoretical analyses. In [ST01], we suggested that smoothed analysis might provide a theoretically analyzable measure of an algorithm's performance that would be more predictive of its behavior in practice. [1].

Algorithms are typically analyzed through either worst-case or average-case complexity. Worst-case analyses may disagree with practical experience because they are dominated by the most pathological input instances. For many algorithms, these pathological inputs are rarely, if ever, encountered in practice, and are only known from lower-bound proofs. In an attempt to create a less pessimistic analysis, researchers introduced average-case analysis, in which one defines a probability distribution on input instances and then measures the expected performance of an algorithm on inputs drawn from that distribution. A low average-case complexity provides some evidence that an algorithm may run quickly in practice. However, this evidence is not conclusive as the inputs encountered by the algorithm in practice may not look like random inputs.

This discrepancy between theoretical and experimental analysis manifests itself in the analysis of linear programming algorithms. The simplex method for linear programming is known to perform very well in practice, but to have exponential worst-case complexity [KM72, Mur80, GS79, Gol83, AC78, Jer73, AZ99]. On the other hand, it is known to have polynomial average-case complexity under a number of notions of average-case [Bor80, Bor77, Sma83, Hai83, AKS87, AM85, Tod86]. Interior-point methods are known to have polynomial worst-case complexity [Kar84]. However, their performance in practice is much better than their worst-case analyses would suggest [IL94, LMS90, EA96]. It has been shown that the average-case complexity of interior-point methods is significantly lower than their worst-case complexity [Ye94, AJPY93, AJPY99]: the term L is replaced by $O(\log n)$.

Smoothed analysis provides an alternative to worst-case and average-case analyses, and also attempts to circumvent the need for a theoretical definition of "practical inputs". The smoothed complexity of an algorithm is defined to be the maximum over its inputs of the expected running time of the algorithm

---

[1] We remark that a similar framework for discrete problems was introduced by Blum and Spencer [BS95]

under slight perturbations of that input. The smoothed complexity is then measured as a function of the input size and the magnitude of the perturbation. While many notions of perturbation are reasonable, most results have been obtained for Gaussian perturbations. The assumption that inputs are subject to perturbation is reasonable in many circumstances: in many real-world numerical and geometric applications, data are derived from experimental and physical measurements and are therefore subject to errors [Wil65, paragraph 2, pp. 62]. Perturbations can also be used to heuristically model the arbitrary decisions that effect to formation of inputs that are presented to algorithms.

Two important aspects of smoothed analysis are:

- Smoothed analysis interpolates between worst-case and average-case analysis: By letting the magnitude of the random perturbation to the data (*e.g.*, the variance of the Gaussian noise) become large, one obtains the traditional average-case complexity measure. By letting the magnitude of the random perturbation go to zero, one obtains the traditional worst-case complexity measure. In between, one obtains a model corresponding to noise in low-order digits of the input.
- The smoothed complexity of an algorithm provides an upper bound on the expected complexity of the algorithm in every neighborhood of inputs. That is, if the smoothed complexity of an algorithm is low, then it will run quickly on inputs drawn from any small neighborhood of inputs.
  Thus, if the inputs presented to an algorithm in practice are subject to perturbation, the smoothed complexity of the algorithm should upper bound the practical performance of the algorithm.

In [ST01], we introduced smoothed complexity by proving that a particular variant of the shadow-vertex simplex method has polynomial smoothed complexity.

1.2. Formal Introduction to Smoothed Analysis

The inputs to most numerical and geometric problems may be viewed as points in a vector space. For example, an m by n real matrix can be viewed as a vector in $\mathbb{R}^{mn}$. Similarly, a set of n points in d dimensions can be viewed as a vector in $\mathbb{R}^{dn}$.

The most natural notion of perturbations of vectors in a real vector space is that of Gaussian perturbations. Recall that a Gaussian random variable with mean 0 and variance σ^2 has density

$$\frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\sigma},$$

and that a Gaussian random vector of variance σ^2 centered at the origin in $\mathbb{R}^n$, denoted $\mathcal{N}(\mathbf{0}, \sigma^2)$, is a vector in which each entry is a Gaussian random variable

of variance σ^2 and mean 0, and has density

$$\frac{e^{-\|\boldsymbol{x}\|^2/2\sigma^2}}{\left(\sqrt{2\pi}\sigma\right)^d}.$$

Definition 1 (Gaussian perturbation). *Let $\hat{\boldsymbol{x}} \in \mathbb{R}^n$. The Gaussian perturbation of $\hat{\boldsymbol{x}}$ of variance σ^2 is the random vector $\boldsymbol{x} = \hat{\boldsymbol{x}} + \boldsymbol{g}$, where $\boldsymbol{g}$ is a Gaussian random vector of variance σ^2, centered at the origin of $\mathbb{R}^n$.*

The Gaussian perturbation of $\hat{\boldsymbol{x}}$ may also be described as a Gaussian random vector of variance σ^2 centered at $\hat{\boldsymbol{x}}$, in which case we observe that $\boldsymbol{x}$ has density $\left(\sqrt{2\pi}\sigma\right)^{-d} e^{-\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2/2\sigma^2}$. In Section 8 we will denote such a density by $\mu_{\boldsymbol{x}}$.

Using the notion of Gaussian perturbation, we define the smoothed value of a function:

Definition 2 (Smoothed value). *Let f be a non-negative function on $\mathbb{R}^n$. The smoothed value of f with respect to Gaussian perturbations of variance σ^2 is given by*

$$\max_{\hat{\boldsymbol{x}}} \mathbf{E}_{\boldsymbol{g}\leftarrow\mathcal{N}(\mathbf{0},\sigma^2)} \left[f(\hat{\boldsymbol{x}} + \|\hat{\boldsymbol{x}}\| \, \boldsymbol{g}) \right]$$

Note that in this definition we multiply the perturbation $\boldsymbol{g}$ by $\|\hat{\boldsymbol{x}}\|$ so that σ represents the magnitude of the perturbation relative to the data.

Definition 3 (Smoothed complexity). *Let A be an algorithm whose inputs can be expressed as vectors in $\mathbb{R}^n$ and let $T_A(\boldsymbol{x})$ be the running time of algorithm A on input $\boldsymbol{x}$. Then the* smoothed complexity *of algorithm A is*

$$\mathcal{C}_A(n,\sigma) = \max_{\hat{\boldsymbol{x}}\in\mathbb{R}^n} \mathbf{E}_{\boldsymbol{g}\leftarrow\mathcal{N}(\mathbf{0},\sigma^2)} \left[T_A(\hat{\boldsymbol{x}} + \|\hat{\boldsymbol{x}}\| \, \boldsymbol{g}) \right].$$

In [ST01], Spielman and Teng consider the complexity of a particular two-phase shadow-vertex simplex method on linear programs of the form

$$\begin{aligned}
&\text{maximize } \boldsymbol{c}^T\boldsymbol{x} \\
&\text{subject to } A\boldsymbol{x} \leq \boldsymbol{b},
\end{aligned} \tag{1}$$

where A is an m-by-n matrix, $\boldsymbol{b}$ is an m-vector, and $\boldsymbol{c}$ is an n-vector. They prove:

Theorem 1.1 (Spielman-Teng). *There is a two-phase shadow-vertex simplex method with time complexity $T(A, \boldsymbol{b}, \boldsymbol{c})$ such that for every m-vector $\boldsymbol{b}$ and n-vector $\boldsymbol{c}$, the smoothed complexity of the algorithm,*

$$\max_{\hat{A}\in\mathbb{R}^{m\times n}} \mathbf{E}_G \left[T\left(\hat{A} + \left\| \hat{A} \right\| G, \boldsymbol{b}, \boldsymbol{c} \right) \right]$$

is polynomial in m, n, and $1/\sigma$, independent of $\boldsymbol{b}$ and $\boldsymbol{c}$, where G is a Gaussian random m by n matrix of variance σ^2 centered at the origin.

One need not limit smoothed analysis to measuring the expected complexity of algorithms in various neighborhoods. It is quite reasonable to prove other facts about the distribution of running times when the expectation does not exist, or when much stronger bounds can be proved. For example, Blum and Dunagan [BD02] prove

Theorem 1.2 (Blum-Dunagan). *Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random vectors in $\mathbb{R}^d$ of variance $\sigma^2 < 1/(2d)$ centered at points each of norm at most 1. There exists a constant c such that the probability that the perceptron algorithm for linear programming takes more than $\frac{cd^3 n^2 \log^2(n/\delta)}{\delta^2 \sigma^2}$ iterations is at most δ.*

2. Notation and Norms

Throughout the paper, we use bold letters such as $\boldsymbol{b}$ and $\boldsymbol{x}$ to denote vectors, capital letters such as A and G to denote matrices, and lower case letters to denote scalars. In any context in which the vector $\boldsymbol{b}$ is present, b_j denotes the jth component of $\boldsymbol{b}$. For a set, N, we let $\boldsymbol{b}_N$ denote the vector obtained by restricting $\boldsymbol{b}$ to the coordinates in N. When indexing and constructing matrices, we use the conventions of Matlab. Thus, $A_{:,B}$ denotes the matrix formed by taking the columns indexed by B, and $A_{N,B}$ denotes the sub-matrix of rows indexed by N and columns indexed by B. For sets, B and N, we let $\bar{B}$ and $\bar{N}$ denote their complements. We also let $\overline{N, B}$ denote the set of pairs $(i, j) \notin (N, B)$; for example, we let $A_{\overline{N,B}}$ denote the set of entries of A not in $A_{N,B}$. For a matrix A and a column vector $\boldsymbol{b}$, we let $[A, \boldsymbol{b}]$ denote the matrix obtained by appending column $\boldsymbol{b}$ to A.

For an event, $\mathcal{E}$, we let $[\mathcal{E}]$ denote the random variable that is 1 when $\mathcal{E}$ is true and is 0 otherwise.

We use of the following vector norms:

- $\|\boldsymbol{x}\| = \sqrt{\sum_i x_i^2}$,
- $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$, and
- $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$,

and note that $\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\| \leq \|\boldsymbol{x}\|_1$.

We also use the following matrix norms:

- $\|A\| = \max_{\boldsymbol{x} \neq 0} \|A\boldsymbol{x}\| / \|\boldsymbol{x}\|$,
- $\|A\|_\infty = \max_{\boldsymbol{x} \neq 0} \|A\boldsymbol{x}\|_\infty / \|\boldsymbol{x}\|_\infty$, and
- $\|A\|_F = \sqrt{\mathbf{trace}(A^T A)}$, the square root of the sum of the squares of entries in A.

We note that

- $\|A\|_\infty = \max_i \|A_{i,:}\|_1$,
- $\|A\|_\infty \leq \sqrt{n} \|A\|$,
- $\|A\| \leq \|A\|_F$, and
- for sets B and N, $\|A_{B,N}\| \leq \|A\|$.

3. Complexity of Linear Programming Algorithms

A linear program is typically specified by a matrix A together with two vectors $\boldsymbol{b}$ and $\boldsymbol{c}$, where where A is an m-by-n matrix, $\boldsymbol{c}$ is an n-dimensional row vector, and $\boldsymbol{b}$ is an m-dimensional column vector. There are several canonical forms of linear programs. For the analyses in this paper, we will consider linear programs of the form

$$\max \boldsymbol{cx} \quad \text{such that } A\boldsymbol{x} \leq \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0},$$

with dual

$$\min \boldsymbol{yb} \quad \text{such that } \boldsymbol{y}A \geq \boldsymbol{c}, \quad \boldsymbol{y} \geq \boldsymbol{0}.$$

We will assume throughout that $m \geq n$.

If they exist, we denote the solutions to the primal and dual by $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$, and note that $\boldsymbol{x}^*$ is an n-dimensional column vector and $\boldsymbol{y}^*$ is an m-dimensional row vector.

A linear programming algorithm should: (1) determine whether or not the linear program is feasible or bounded; and, (2) if the program is feasible and bounded, output a solution. One can either insist that the solution be a precisely optimal solution to the linear program, or merely a feasible point at which the objective function is approximately optimized.

The best bounds on the worst-case complexity of interior-point methods, and for linear programming in general, were first obtained by Gonzaga [Gon88] and Vaidya [Vai90], who showed how to solve linear programs in $O(m^3 L)$ arithmetic operations[2], where $m \geq n$ and L is a parameter measuring the precision needed to perform the arithmetic operations exactly, and which here also appears in the number of arithmetic operations performed. The definition of L varies in the literature: Khachiyan [Kha79], Karmarkar [Kar84], and Vaidya [Vai90] define L for integer matrices A to be some constant times

log(largest absolute value of the determinant of any square sub-matrix of A)
$$+ \log(\|\boldsymbol{c}\|_\infty) + \log(\|\boldsymbol{b}\|_\infty) + \log(m + n).$$

Under this definition, L is not efficiently computable, and unless A comes from a very special class of matrices, it is difficult to find L below $\Omega(n)$. Others use cruder bounds such as the total number of bits in a row of the matrix or the total number of bits in the entire matrix [Wri96].

---

[2] Vaidya's algorithm is somewhat faster as its complexity is $O((m + n)n^2 + (m + n)^{1.5}n)L$

To understand the time complexity of interior-point algorithms, we note that they are typically divided into three phases:

[Initialization]: In this phase, the algorithm determines whether or not the program is feasible and bounded; and, if it is feasible and bounded, returns a feasible point.

[Iteration]: In this phase, the algorithm iteratively finds feasible points on which the objective function becomes increasingly better.

[Termination]: In this phase, the algorithm jumps from a feasible point that is close to optimal to the exact optimal solution of the linear program.

Of course, if one merely desires an approximate solution to the linear program, then one can skip the termination phase. However, the dependency on L appears in both the initialization and termination phase. So, the worst-case complexity of linear programming algorithms is not decreased by merely asking for an approximate solution.

The kernel of an interior-point algorithm is the iteration phase, during which feasible points of increasing quality are computed. A typical measure of quality in a primal algorithm is the *optimality gap* between the objective function at the current point and the optimal, while in a primal-dual algorithm it is the *duality gap* between the current primal and dual feasible points. In either case, one can prove that after k iterations the gap decreases by the multiplicative factor $\left(1 - \frac{c}{\sqrt{m}}\right)^k$, for some constant c [Ren88, Vai90, Ye97]. If performed carefully, each of these iterations has complexity $O(m^{5/2})$ [Gon88]. Therefore, the total number of arithmetic operations required to reduce the gap from R to ϵ is $O(m^3 \log(R/\epsilon))$. The worst-case complexity bounds come from the facts that a typical interior-point algorithm discovers a feasible point with initial gap bounded by $R = 2^{O(L)}$ in the initialization phase, and requires a point with gap less than $\epsilon = 2^{-O(L)}$ to start the termination phase.

In practice, the speed of interior-point methods is much better than that proved in their worst-case analyses [IL94, LMS90, EA96]. This difference in speed seems to have two sources: first, the upper bound of L is overly pessimistic; and, second, the improvement made at each iteration is typically much better than $\left(1 - \frac{c}{\sqrt{m}}\right)$. However, we note that Todd [Tod94] and Todd and Ye [TY96] have exhibited linear programs in which $\Omega(n^{1/3})$ iterations are required to improve the gap by a constant factor.

The average case analyses of interior-point methods on random inputs proceed by upper bounding R for the initial point constructed by an algorithm and by lower bounding the ϵ at which the termination phase can succeed. For example, Ye [Ye94] presents an algorithm that takes $O(m^{1/2} \log(\Sigma/\xi))$ iterations, where $\Sigma \stackrel{\text{def}}{=} \max(\|\boldsymbol{x}^*\|, \|\boldsymbol{y}^*\|, \|\boldsymbol{b} - A\boldsymbol{x}^*\|, \|\boldsymbol{c} - \boldsymbol{y}^*A\|)$, and ξ is a condition number measuring the smallest non-zero slack at the solutions.

Our smoothed analysis of interior-point methods follows a similar approach. We apply the bounds on the smoothed value of Renegar's condition number from [DST02] to bound R for the initial point provided by Renegar's algorithm [Ren95b] (or Σ for Ye's algorithm [Ye94] via Lemma 6.7). We then define a termination condition number, δ, that measures how close to optimal a feasible point $\boldsymbol{x}$ needs to be before the termination phase will round it to the optimal solution. In Section 6, we define quantities $\alpha_P, \beta_P, \alpha_D, \beta_P$, and, following Ye, define ξ to be $\min(\alpha_P, \beta_P, \alpha_D, \beta_P)$. We then prove that our condition number, δ, is bounded by a function of ξ and the quantity γ, also defined in Section 6. Relations between our δ and Ye's ξ include:

- ξ provides a bound on how small the duality gap of a primal-dual pair must be before Ye's finite termination algorithm [Ye92] may be applied, while
- δ provides a bound on how close to optimal a primal feasible point must be to optimal before the termiation algorithm of Section 5 may be applied, and
- we only define δ for programs with unique primal and dual optimal solutions, as all programs considered in this paper have this property with probability 1. However, we expect the definition could be extended to degenerate linear programs.

By combining the smoothed analyses of Renegar's condition number and δ, we show that the smoothed complexities of Renegar's Algorithm [Ren95b] and Ye's Algorithm [Ye94] are $O(m^3 \log(m/\sigma))$. Essentially, this analysis replaces the dependency on L in the initialization and termination phases with a dependency on $\log(m/\sigma)$. We conjecture that one can improve this smoothed complexity estimate by proving that the smoothed number of iterations taken by an interior-point method is less than $O(\sqrt{m} \log(m/\sigma))$.

Renegar [Ren95b, Ren95a, Ren94] defined a condition number $C(A, \boldsymbol{b}, \boldsymbol{c})$ of a linear program, and developed an algorithm for the initialization of an interior-point method that runs in time $O(m^3 \log(C(A, \boldsymbol{b}, \boldsymbol{c})))$ and returns a feasible point with initial optimality gap $R \le O(mC(A, \boldsymbol{b}, \boldsymbol{c}))$. Applying a primal iteration phase to this feasible point, one obtains an algorithm that after $O(\sqrt{m} \log(\epsilon^{-1}mC(A, \boldsymbol{b}, \boldsymbol{c})))$ rounds and $O(m^3 \log(\epsilon^{-1}mC(A, \boldsymbol{b}, \boldsymbol{c})))$ arithmetic operations produces points with optimality gap at most ϵ. Renegar's condition number will be discussed further in Section 4.

Dunagan, Spielman and Teng [DST02] perform a smoothed analysis of Renegar's condition number and prove:

Theorem 3.1 (Dunagan-Spielman-Teng). *Let $\hat{A}$ be an m-by-n matrix for* $m \ge n$, $\hat{\boldsymbol{b}}$ *an m-vector, and $\hat{\boldsymbol{c}}$ an n-vector for which $\left\|\hat{A}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\right\|_F \le 1$. Let A, $\boldsymbol{b}$ and $\boldsymbol{c}$ be the Gaussian perturbations of $\hat{A}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{c}}$ of variance $\sigma \le 1/\sqrt{mn}$. Then,*

$$\mathbf{E}_{A,\boldsymbol{b},\boldsymbol{c}} \left[\log(C(A, \boldsymbol{b}, \boldsymbol{c}))\right] \le O(\log(m/\sigma)).$$

This theorem has two roles in this paper: it is used in the proof that it is unlikely that δ is small, and it is used to bound the quality of the initial point returned by Renegar's algorithm [Ren95b]. Combining this theorem with the analysis of Renegar's algorithm [Ren95b], we find that the smoothed complexity of finding an ϵ-optimal solution to a linear program with Renegar's algorithm is $O(m^3 \log(m/\sigma\epsilon))$. Similarly, one can apply Lemma 6.7 and Proposition 6.1 to this bound to bound Σ and show that Ye's algorithm has a similar complexity. However, we suspect that a more direct analysis of Σ might be possible.

In Section 5, we define a simple termination algorithm that takes $O(m^3)$ arithmetic operations. We define $\delta(A, \boldsymbol{b}, \boldsymbol{c})$ to be the greatest number such that $\boldsymbol{c}\boldsymbol{x}^* - \boldsymbol{c}\boldsymbol{x} \leq \delta(A, \boldsymbol{b}, \boldsymbol{c})$ implies that the termination algorithm is successful. Thus, after $O(\sqrt{m} \log(mC(A, \boldsymbol{b}, \boldsymbol{c})\delta(A, \boldsymbol{b}, \boldsymbol{c})^{-1}))$ interior-point iterations, and $O(m^3 \log(mC(A, \boldsymbol{b}, \boldsymbol{c})\delta(A, \boldsymbol{b}, \boldsymbol{c})^{-1}))$ arithmetic operations, one can apply the termination phase to find the exact solution to the linear program. Like Karmarkar [Kar84], we handle the technical difficulty that the algorithm does not know $C(A, \boldsymbol{b}, \boldsymbol{c})$ or $\delta(A, \boldsymbol{b}, \boldsymbol{c})$ by periodically attempting to terminate, but only once every $\sqrt{n}$ iterations so as not to increase the complexity of the algorithm.

In Theorem 5.2, proved over Sections 6, 7 and 8, we show that the smoothed value of $\max\left(1, \log\left(\delta(A, \boldsymbol{b}, \boldsymbol{c})^{-1}\right)\right)$ is $O(\log(m/\sigma))$. We note that as the termination phase only requires a good primal solution, this analysis can be applied to a wide variety of linear programming algorithms, including Vaidya's [Vai90], the ellipsoid algorithm, and the recent algorithm of Bertsimas and Vempala [BV02]. For interior-point algorithms, we prove:

Theorem 3.2 (Smoothed Complexity of IPM). *Let $\hat{A}$ be an m-by-n matrix with $m \geq n$, $\hat{\boldsymbol{b}}$ be an m-vector, and $\hat{\boldsymbol{c}}$ be an n-vector for which $\left\|\hat{A}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\right\|_F \leq 1$. Let A, $\boldsymbol{b}$ and $\boldsymbol{c}$ be the Gaussian perturbations of $\hat{A}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{c}}$ of variance $\sigma < 1/\sqrt{mn}$. Let $T(A, \boldsymbol{b}, \boldsymbol{c})$ denote the complexity of Renegar's interior-point algorithm with the periodic application of the termination procedure described in Section 5. Then,*

$$\mathbf{E}_{A, \boldsymbol{b}, \boldsymbol{c}}\left[T(A, \boldsymbol{b}, \boldsymbol{c})\right] \leq O(m^3 \log(m/\sigma)).$$

A similar statement may be proved for Ye's algorithm [Ye94].

While this is the statement of the complexity that is most natural for our proof techniques, we note that it is not exactly the form specified in Definition 3. The difference comes from the upper bounds on σ and $\left\|\hat{A}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\right\|_F$ in the statement of the theorem. As the behavior of the interior-point methods are unchanged by multiplicative changes to $\hat{A}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{c}}$, only the upper bound on σ is significant: if $\left\|\hat{A}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\right\|_F \geq 1$, then one can scale down $\hat{A}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{c}}$, and σ to make $\left\|\hat{A}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\right\|_F = 1$. One could adjust Theorem 3.2 in two ways to handle $\sigma > 1/\sqrt{mn}$: one could either extend the proofs, or one could use Theorem 3.2 as a black-box and derive the more general statement from it. Such a proof could proceed by observing

that a Gaussian of variance σ^2 is the sum of a Gaussian of variance τ^2 and a Gaussian of variance $\sigma^2 - \tau^2$. Thus, one can apply Theorem 3.2 with a Gaussian of variance τ^2 to the result of perturbing the original data by a Gaussian of variance $\sigma^2 - \tau^2$, for an appropriate choice of τ. The reader can find a precise implementation of this technique in [ST01, Section 5.1].

4. Renegar's Condition Number for Linear Programming

In an effort to develop a parameter in which to measure the complexity of linear programming that was more natural than L, Renegar [Ren95b, Ren95a, Ren94], introduced the condition number, $C(A, \boldsymbol{b}, \boldsymbol{c})$, of a linear program and developed an interior-point method that runs in time $O(m^3 \log(C(A, \boldsymbol{b}, \boldsymbol{c})\epsilon^{-1}))$. In contrast with the parameter L, $C(A, \boldsymbol{b}, \boldsymbol{c})$ is naturally defined for rational or real matrices A. Moreover, $C(A, \boldsymbol{b}, \boldsymbol{c})$ is often much smaller than L.

Formally, we define the distance of a linear program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$ to primal ill-posedness to be

$$\kappa_P(A, \boldsymbol{b}) = \begin{cases} \sup\left\{\kappa : \left\|A - A', \boldsymbol{b} - \boldsymbol{b}'\right\|_F \leq \kappa \text{ implies } A'\boldsymbol{x} \leq \boldsymbol{b}', \boldsymbol{x} \geq 0 \text{ is feasible}\right\} \\ \qquad \text{if } A\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq 0 \text{ is feasible, and} \\ \sup\left\{\kappa : \left\|A - A', \boldsymbol{b} - \boldsymbol{b}'\right\|_F \leq \kappa \text{ implies } A'\boldsymbol{x} \leq \boldsymbol{b}', \boldsymbol{x} \geq 0 \text{ is infeasible}\right\} \\ \qquad \text{if } A\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq 0 \text{ is infeasible.} \end{cases}$$

The distance to dual ill-posedness, $\kappa_D(A, \boldsymbol{c})$, is defined similarly. We then define $C(A, \boldsymbol{b}, \boldsymbol{c})$ to be the maximum of the primal condition number $C_P(A, \boldsymbol{b})$ and the dual condition number $C_D(A, \boldsymbol{c})$, where $C_P(A, \boldsymbol{b})$ and $C_D(A, \boldsymbol{c})$ are the normalized distances to primal and dual ill-posedness:

$$C_P(A, \boldsymbol{b}) = \left\|A, \boldsymbol{b}\right\|_F / \kappa_P(A, \boldsymbol{b}) \text{ and } C_D(A, \boldsymbol{c}) = \left\|A, \boldsymbol{c}\right\|_F / \kappa_D(A, \boldsymbol{c}).$$

We remark that, with this normalization, C_P and C_D are always at least 1.

We also note that the linear programs for which Todd [Tod94] and Todd and Ye [TY96] prove a $\Omega(n^{1/3})$ iteration lower bound have exponentially poor condition. It is not known if one can prove such an iteration lower bound for a well-conditioned linear program.

5. Termination

One can often terminate algorithms that approach the optimal solution of a linear program by using a good solution to guess the optimal solution. The process by which this is done is often called termination or rounding. Termination is possible because at the optimal solution a number of the inequalities are tight, and the knowledge of the identity of these inequalities is enough to reconstruct the optimal solution. Thus, most termination algorithms work by guessing that

the inequalities having the least slack at a very good solution are those which have no slack at the optimal solution.

We being by recalling the facts that we will use to prove that termination is possible, ignoring complications that may occur with probability zero for perturbed A, $\boldsymbol{b}$ and $\boldsymbol{c}$. We begin with

Proposition 5.1. *For Gaussian distributed A, $\boldsymbol{b}$ and $\boldsymbol{c}$, with probability 1, the program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$ is either infeasible, unbounded, or has unique primal and dual optimal solutions, $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$. Moreover, $\boldsymbol{x}^*$ makes tight exactly n of the inequalities $\{x_i \geq 0\} \cup \{A_{j,:}\boldsymbol{x} \leq b_j\}$ and $\boldsymbol{y}^*$ makes tight exactly m of the inequalities $\{y_j \geq 0\} \cup \{\boldsymbol{y}A_{:,i} \geq c_i\}$.*

Proof. If the primal program is feasible and bounded but does not have a unique optimal solution, then the space of optimal solutions must lie in a subspace defined by fewer than n of the inequalities $\{x_i = 0\} \cup \{A_{j,:}\boldsymbol{x} = b_j\}$. As all the points in this space must be optimal, $\boldsymbol{c}$ must be orthogonal to the subspace. However, as this restricts $\boldsymbol{c}$ to a set of measure zero and the number of such possible subspaces is finite given A and $\boldsymbol{b}$, this is an event with probability zero. By symmetry, the same holds for the optimal solution of the dual program. To prove the second part, we note that if $n + 1$ of the inequalities are tight at $\boldsymbol{x}^*$, then these inequalities form a system of $n+1$ equations in n variables that has a solution. As any such degeneracy has probability zero, and there are only finitely many such possible degeneracies, the probability of this happening is zero.

We now recall the Duality Theorem of Linear Programming:

Theorem 5.1 (LP duality). *For a linear program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$,*

> *(Weak Duality) for every primal feasible $\boldsymbol{x}$ and dual feasible $\boldsymbol{y}$, $\boldsymbol{y}\boldsymbol{b} \geq \boldsymbol{c}\boldsymbol{x}$, and*
> *(Strong Duality) if the linear program is bounded and feasible, then for primal optimal $\boldsymbol{x}^*$ and a dual optimal $\boldsymbol{y}^*$ we have $\boldsymbol{y}^*\boldsymbol{b} = \boldsymbol{y}^*A\boldsymbol{x}^* = \boldsymbol{c}\boldsymbol{x}^*$.*

For a feasible and bounded linear program $(A, \boldsymbol{b}, \boldsymbol{c})$ with unique optimal primal and dual solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$, we define

$$B = \{i : \boldsymbol{x}_i^* > 0\}$$
$$N = \{j : \boldsymbol{y}_j^* > 0\},$$

and we say that the program is of type (B, N). If the primal and dual optimal solutions are unique, then the strict complementarity theorem (see [Ye97, Theorem 1.14]) implies that N and B also index the tight constraints:

$$N = \{j : A_{j,:}\boldsymbol{x}^* = \boldsymbol{b}_j\}, \text{ and}$$
$$B = \{i : \boldsymbol{y}^*A_{:,i} = \boldsymbol{c}_i\}.$$

We will consider the following termination scheme: suppose $\boldsymbol{x}$ is an approximate solution to the primal program, and let $B(\boldsymbol{x})$ and $N(\boldsymbol{x})$ be the set of indices such that
$$\{\boldsymbol{x}_i : i \notin B(\boldsymbol{x})\} \cup \{\boldsymbol{b}_j - A_{j,:}\boldsymbol{x} : j \in N(\boldsymbol{x})\}$$
be the smallest n values in $\{\boldsymbol{x}_i\} \cup \{\boldsymbol{b}_j - A_{j,:}\boldsymbol{x}\}$. We then guess that the optimal solution is the solution to the following linear system: $\boldsymbol{x}_i = 0$ for $i \notin B(\boldsymbol{x})$ and $\boldsymbol{b}_j - A_{j,:}\boldsymbol{x} = 0$ for $j \in N(\boldsymbol{x})$. We will show that if $\boldsymbol{x}$ is sufficiently close to optimal, then this guess is correct. We now define $\delta(A, \boldsymbol{b}, \boldsymbol{c})$ to measure how close to optimal $\boldsymbol{x}$ needs to be.

Definition 4 ($\delta(A, \boldsymbol{b}, \boldsymbol{c})$). *For a feasible and bounded linear program specified by A, $\boldsymbol{b}$ and $\boldsymbol{c}$, we define*
$$\delta(A, \boldsymbol{b}, \boldsymbol{c})$$
to be the supremum of the δ for which
$$(\boldsymbol{c}\boldsymbol{x}^* - \boldsymbol{c}\boldsymbol{x}) < \delta \text{ implies } B(\boldsymbol{x}) = B \text{ and } N(\boldsymbol{x}) = N. \tag{2}$$
For unbounded or infeasible programs, we set δ to ∞.

The main technical contribution of this paper is:

Theorem 5.2. *Let $\hat{A}$ be an m-by-n matrix , with $m \geq n$, $\hat{\boldsymbol{b}}$ be an m-vector, and $\hat{\boldsymbol{c}}$ be an n-vector for which $\left\|\hat{A}\right\|, \left\|\hat{\boldsymbol{b}}\right\|, \|\hat{\boldsymbol{c}}\| \leq 1$. Let A, $\boldsymbol{b}$ and $\boldsymbol{c}$ be a Gaussian random matrix and two Gaussian random vectors of variance σ^2 centered at $\hat{A}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{c}}$, respectively. Then, for $\sigma^2 \leq 1$,*
$$\mathbf{E}\left[\max\left(1, \log\left(\frac{1}{\delta(A, \boldsymbol{b}, \boldsymbol{c})}\right)\right)\right] \leq O(\log(m\sigma^{-1})).$$

Our proof of Theorem 5.2 is broken into three sections. In Section 6 we define the geometric quantities that we will use to bound $\delta(A, \boldsymbol{b}, \boldsymbol{c})$, and state the a relation between $\delta(A, \boldsymbol{b}, \boldsymbol{c})$ and these quantities (Lemma 6.5). This relation is proved in Section 7. In Section 8, we prove bounds on the probabilities that these geometric quantities are small. The rest of the material in Section 6 is a routine calculation using the results of Sections 7, 8 and Theorem 3.1. The reader will probably be most interested in Section 8, which we begin with an intuitive explanation of how the probability estimates are obtained, carefully explain the tools used to make these arguments rigorous, and then finally apply these tools to obtain the probability bounds.

We remark that Theorem 5.2 depends very little on the properties of Gaussian random variables. Aside for the bound on $\mathbf{E}\left[\log \|A\|\right]$ of Proposition 6.1, which is easily generalized to other distributions, the only facts about Gaussian random variables used are that proved in Lemma 8.1, and those used in the proof of Theorem 3.1. Thus, one could prove statements similar to Theorem 5.2 for a number of families of perturbations. For example, it would be easy to adjust the proof techniques to handle perturbations by matrices and vectors of independent uniform random variables.

6. Smoothed Analysis of $\log(1/\delta)$

Our analysis of the probability that $\delta(A, \boldsymbol{b}, \boldsymbol{c})$ is small will be divided into two parts: a geometric condition for $\delta(A, \boldsymbol{b}, \boldsymbol{c})$ to be small, and a bound on the probability that this geometric condition is satisfied.

To described the geometric condition, we define the following five quantities for bounded and feasible linear programs with unique optimal primal and dual solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$.

$$
\begin{aligned}
\alpha_P(A, \boldsymbol{b}, \boldsymbol{c}) &\stackrel{\text{def}}{=} \min_{i \in B} x_i^*, \\
\alpha_D(A, \boldsymbol{b}, \boldsymbol{c}) &\stackrel{\text{def}}{=} \min_{j \in N} y_j^*, \\
\beta_P(A, \boldsymbol{b}, \boldsymbol{c}) &\stackrel{\text{def}}{=} \min_{j \in \bar{N}} b_j - A_{j,:}\boldsymbol{x}^*, \\
\beta_D(A, \boldsymbol{b}, \boldsymbol{c}) &\stackrel{\text{def}}{=} \min_{j \in \bar{B}} \boldsymbol{y}^* A_{:,i} - c_i, \\
\gamma(A, \boldsymbol{b}, \boldsymbol{c}) &\stackrel{\text{def}}{=} \min_{k \in B} \mathbf{dist}\left(A_{N,k}, \mathbf{Span}\left(A_{N,B-k}\right)\right).
\end{aligned}
$$

The geometric condition is that one of these five quantities is small.

When A, $\boldsymbol{b}$ and $\boldsymbol{c}$ are clear from context, we will just write α_P, α_D, β_D, β_P, γ or δ. Note that

$$
\alpha_D(A, \boldsymbol{b}, \boldsymbol{c}) = \alpha_P(-A^T, -\boldsymbol{c}^T, -\boldsymbol{b}^T) \quad \text{and} \quad \beta_D(A, \boldsymbol{b}, \boldsymbol{c}) = \beta_P(-A^T, -\boldsymbol{c}^T, -\boldsymbol{b}^T).
$$

In Section 7, we prove

Lemma 6.1. *For a linear program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$ with unique optimal primal and dual solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$, let*

$$
\xi(A, \boldsymbol{b}, \boldsymbol{c}) \stackrel{\text{def}}{=} \min\left(\alpha_P(A, \boldsymbol{b}, \boldsymbol{c}), \alpha_D(A, \boldsymbol{b}, \boldsymbol{c}), \beta_P(A, \boldsymbol{b}, \boldsymbol{c}), \beta_D(A, \boldsymbol{b}, \boldsymbol{c})\right).
$$

Then,

$$
\delta(A, \boldsymbol{b}, \boldsymbol{c}) \geq \frac{\xi(A, \boldsymbol{b}, \boldsymbol{c})^2 \gamma(A, \boldsymbol{b}, \boldsymbol{c})}{2 \max(1, \sqrt{n}\, \|A\|)\, (1 + \|A\|)}.
$$

We remark that Ye [Ye94] shows how to terminate given a pair with duality gap less than $O(\xi^2/n)$. Thus, some simplification of our analysis can be obtained by restricting one's attention to primal-dual algorithms, as it becomes unnecessary to bound the probability that γ is small.

We define $\mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})$ to be the event that the linear program specified by A, $\boldsymbol{b}$, $\boldsymbol{c}$ is feasible and bounded. In Section 8, we prove

Lemma 6.2 (Probability of small α). *Under the conditions of Theorem 5.2,*

$$
\Pr\left[\alpha_P(A, \boldsymbol{b}, \boldsymbol{c}) \leq \frac{\epsilon}{(\|A\| + 2)^2\,(\|\boldsymbol{x}^*\| + 1)} \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right] \leq \frac{8\epsilon n(m+1)}{\sigma^2}.
$$

Lemma 6.3 (Probability of small β). *Under the conditions of Theorem 5.2,*

$$\Pr\left[\beta_P(A,\boldsymbol{b},\boldsymbol{c}) \leq \frac{\epsilon}{\max\left(1, \|A\|\, \|\boldsymbol{x}^*\|\right)}\ \text{ and }\mathcal{F}(A,\boldsymbol{b},\boldsymbol{c}))\right] \leq \frac{4\epsilon m}{\sigma^2}.$$

Lemma 6.4 (Probability of small γ). *Under the conditions of Theorem 5.2,*

$$\Pr_{A_{N,B}}\left[\gamma(A,\boldsymbol{b},\boldsymbol{c}) \leq \frac{\epsilon}{\left(1 + \|\boldsymbol{x}^*\|^2 + \|\boldsymbol{y}^*\|^2\right)\left(\|A\|+3\right)}\ \text{ and }\mathcal{F}(A,\boldsymbol{b},\boldsymbol{c})\right] \leq \frac{3e\epsilon n}{\sigma^2}.$$

Using these three lemmas, we can reduce our analysis of the probability that $\delta(A,\boldsymbol{b},\boldsymbol{c})$ is small to an analysis of the probability that $\|\boldsymbol{x}^*\|$, $\|\boldsymbol{y}^*\|$ or $\|A\|$ is large.

Lemma 6.5. *Under the conditions of Theorem 5.2,*

$$\Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\delta(A,\boldsymbol{b},\boldsymbol{c})\left(\|A\|+3\right)^7\left(1+\|\boldsymbol{x}^*\|+\|\boldsymbol{y}^*\|\right)^4 \leq \epsilon\right] \leq \frac{22\epsilon^{1/3}n^{1/6}(n+1)(m+1)}{\sigma^2}. \tag{3}$$

Proof. As $\delta(A,\boldsymbol{b},\boldsymbol{c})$ is infinite for infeasible or unbounded programs,

$$\Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\delta(A,\boldsymbol{b},\boldsymbol{c})\left(\|A\|+3\right)^7\left(1+\|\boldsymbol{x}^*\|+\|\boldsymbol{y}^*\|\right)^4 \leq \epsilon\right]$$
$$= \Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\delta(A,\boldsymbol{b},\boldsymbol{c})\left(\|A\|+3\right)^7\left(1+\|\boldsymbol{x}^*\|+\|\boldsymbol{y}^*\|\right)^4 \leq \epsilon \text{ and } \mathcal{F}(A,\boldsymbol{b},\boldsymbol{c})\right].$$

As $\alpha_D(A,\boldsymbol{b},\boldsymbol{c}) = \alpha_P(-A^t, -\boldsymbol{c}^t, -\boldsymbol{b}^t)$, and

$$\frac{\epsilon}{\left(\|A\|+2\right)^2\left(\|\boldsymbol{x}^*\|+1\right)} \leq \frac{\epsilon}{\max\left(1, \|A\|\,\|\boldsymbol{x}^*\|\right)},$$

Lemmas 6.2 and 6.3 imply

$$\Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\min\left(\alpha_P, \alpha_D, \beta_P, \beta_D\right) \leq \frac{\epsilon}{\left(\|A\|+2\right)^2\left(\max\left(\|\boldsymbol{x}^*\|, \|\boldsymbol{y}^*\|\right)+1\right)}\ \text{ and } \mathcal{F}(A,\boldsymbol{b},\boldsymbol{c})\right]$$
$$\leq \frac{8\epsilon n(m+1) + 8\epsilon m(n+1) + 4\epsilon m + 4\epsilon n}{\sigma^2}$$
$$\leq \frac{8\epsilon\left((n+1/2)(m+1) + (m+1/2)(n+1)\right)}{\sigma^2}. \tag{4}$$

Let $\xi = \min\left(\alpha_P, \alpha_D, \beta_P, \beta_D\right)$. If

$$\xi^2\gamma \leq \frac{\epsilon^3}{\left(\|A\|+3\right)^5\left(1+\|\boldsymbol{x}^*\|+\|\boldsymbol{y}^*\|\right)^4}$$
$$\leq \left(\frac{\epsilon}{\left(\|A\|+2\right)^2\left(\max\left(\|\boldsymbol{x}^*\|, \|\boldsymbol{y}^*\|\right)+1\right)}\right)^2\left(\frac{\epsilon}{\left(1+\|\boldsymbol{x}^*\|^2+\|\boldsymbol{y}^*\|^2\right)\left(\|A\|+3\right)}\right)$$

then either

$$\xi \leq \frac{\epsilon}{\left(\|A\| + 2\right)^2 \left(\max\left(\|\boldsymbol{x}^*\|, \|\boldsymbol{y}^*\|\right) + 1\right)}$$

or

$$\gamma \leq \frac{\epsilon}{\left(1 + \|\boldsymbol{x}^*\|^2 + \|\boldsymbol{y}^*\|^2\right)\left(\|A\| + 3\right)}.$$

So, Lemma 6.4 and inequality (4) imply

$$\Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\xi^2\gamma \leq \frac{\epsilon^3}{\left(\|A\| + 3\right)^5\left(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|\right)^4} \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$

$$\leq \frac{3e\epsilon n}{\sigma^2} + \frac{8\epsilon\left((n + 1/2)(m + 1) + (m + 1/2)(n + 1)\right)}{\sigma^2}$$

$$\leq \frac{17\epsilon(n + 1)(m + 1)}{\sigma^2}, \qquad\qquad\qquad \text{as } 3en \leq 5(n + m).$$

As Lemma 6.1 tells us that

$$\delta(A, \boldsymbol{b}, \boldsymbol{c}) \geq \frac{\xi^2\gamma}{2\max(1, \sqrt{n}\,\|A\|)\left(1 + \|A\|\right)},$$

we obtain

$$\Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\delta \leq \frac{\epsilon^3}{\left(\|A\| + 3\right)^5\left(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|\right)^4}\left(\frac{1}{2\max(1, \sqrt{n}\|A\|)(1 + \|A\|)}\right) \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$

$$\leq \Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\xi^2\gamma \leq \frac{\epsilon^3}{\left(\|A\| + 3\right)^5\left(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|\right)^4} \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$

$$\leq \frac{17\epsilon(n + 1)(m + 1)}{\sigma^2}.$$

From this inequality, we derive

$$\Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\delta\left(\|A\| + 3\right)^7\left(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|\right)^4 \leq \frac{\epsilon^3}{2\sqrt{n}} \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$

$$\leq \Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\begin{matrix}\delta\left(\|A\| + 3\right)^5\left(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|\right)^4\max(1, \|A\|)\left(1 + \|A\|\right) \leq \frac{\epsilon^3}{2\sqrt{n}} \\ \text{and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\end{matrix}\right]$$

$$\leq \frac{17\epsilon(n + 1)(m + 1)}{\sigma^2}.$$

The lemma now follows by changing $\epsilon^3/(2\sqrt{n})$ to ϵ and observing $17 \cdot 2^{1/3} \leq 22$.

To convert this bound on the probability that δ is small to a bound on the expectation of $\log(1/\delta)$, we will use the following technical lemma:

Lemma 6.6. *Let x be a non-negative random variable for which there exist constants α and k satisfying $\log(\alpha)/k \geq 1$ such that*

$$\Pr\left[x \leq \epsilon\right] \leq \alpha\epsilon^k.$$

Then,

$$\mathbf{E}\left[\max\left(1, \log(1/x)\right)\right] \leq \frac{1 + \log\alpha}{k}.$$

Proof. We compute

$$
\begin{aligned}
\mathbf{E}\left[\max\left(1, \log(1/x)\right)\right] &= \int_{t=0}^{\infty} \Pr\left[\max\left(1, \log(1/x)\right) \geq t\right] dt \\
&= \int_{t=0}^{1} dt + \int_{t=1}^{\infty} \Pr\left[\log(1/x) \geq t\right] dt \\
&\leq \int_{t=0}^{\frac{\log\alpha}{k}} dt + \int_{\frac{\log\alpha}{k}}^{\infty} \Pr\left[\log(1/x) \geq t\right] dt, \quad \text{as } (\log\alpha)/k \geq 1, \\
&\leq \int_{t=0}^{\frac{\log\alpha}{k}} dt + \int_{\frac{\log\alpha}{k}}^{\infty} \alpha e^{-tk} dt \\
&= \frac{\log\alpha}{k} + \frac{1}{k}.
\end{aligned}
$$

From this, we obtain

Corollary 6.1. *Under the conditions of Theorem 5.2,*

$$
\mathbf{E}\left[\max\left(1, \log\left(\frac{1}{\delta(A, \boldsymbol{b}, \boldsymbol{c})}\right)\right)\right]
$$
$$
\leq 3(\log(22\sigma^{-2}(m+1)^{13/6}) + 1) + 7\log(\|A\| + 3) + 4\log(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|)
$$

Proof. Applying Lemma 6.6 to (3), and recalling $m \geq n$, we obtain

$$
\mathbf{E}\left[\max\left(1, \log\left(\frac{1}{\delta(A, \boldsymbol{b}, \boldsymbol{c})}\right)\right)\right] - 7\mathbf{E}\left[\log(\|A\| + 3)\right] - 4\mathbf{E}\left[\log(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|)\right]
$$
$$
\leq \mathbf{E}\left[\max\left(1, \log\left(\frac{1}{\delta(A, \boldsymbol{b}, \boldsymbol{c})\left(\|A\| + 3\right)^7 \left(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|\right)^4}\right)\right)\right]
$$
$$
\leq 3\left(\log(22(m+1)^{13/6}/\sigma^2) + 1\right).
$$

To bound $\mathbf{E}\left[\log(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|)\right]$, we note that Renegar [Ren95a, Propositions 2.2 and 2.3] has proved

Lemma 6.7 (Norms of optimal solutions).

$$\max\left(\|\boldsymbol{x}^*\|, \|\boldsymbol{y}^*\|\right) \leq C(A, \boldsymbol{b}, \boldsymbol{c})^2.$$

So, we may apply Theorem 3.1 to bound the norms of $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$. To bound the norm of A, we apply:

Proposition 6.1. *Let A be a Gaussian perturbation of variance $\sigma^2 \leq 1$ of an m-by-n matrix $\hat{A}$ of norm at most 1. Then,*

$$\mathbf{E}\left[\log(\|A\| + 3)\right] \leq \log\left(\left(\sqrt{n} + \sqrt{m}\right)\sigma + 4\right)$$

Proof. Write $A = \hat{A} + \sigma G$ where G is a Gaussian random matrix of variance 1 centered at the origin and $\left\|\hat{A}\right\| \leq 1$. Seginer [Seg00] proves that $\mathbf{E}\left[\|G\|\right] \leq \sqrt{n} + \sqrt{m}$, which implies $\mathbf{E}\left[\|\sigma G\|\right] \leq \sigma(\sqrt{n} + \sqrt{m})$ and

$$\mathbf{E}\left[\|A\| + 3\right] \leq \mathbf{E}\left[\left\|\hat{A}\right\| + \|\sigma G\| + 3\right] \leq \left(\sqrt{n} + \sqrt{m}\right)\sigma + 4.$$

As the logarithm is a convex function,

$$\mathbf{E}\left[\log(\|A\| + 3)\right] \leq \log\left(\mathbf{E}\left[\|A\| + 3\right]\right) \leq \log\left(\left(\sqrt{n} + \sqrt{m}\right)\sigma + 4\right).$$

Combining these estimates on the probability that $\|A\|$, $\|\boldsymbol{x}^*\|$ or $\|\boldsymbol{y}^*\|$ is large with Corollary 6.1, we prove the main theorem:

Proof (of Theorem 5.2). To bound the terms obtained in Corollary 6.1, we first apply Proposition 6.1 to obtain

$$\mathbf{E}\left[\log(\|A\| + 3)\right] \leq \log(\sqrt{n} + \sqrt{m} + 4).$$

We then apply Lemma 6.7 to show

$$(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|) \leq 3C(A, \boldsymbol{b}, \boldsymbol{c})^2,$$

and Theorem 3.1 to obtain

$$
\begin{aligned}
\mathbf{E}\left[\log(1 + \|\boldsymbol{x}^*\| + \|\boldsymbol{y}^*\|)\right] &\leq \mathbf{E}\left[\log\left(3C(A, \boldsymbol{b}, \boldsymbol{c})^2\right)\right] \\
&\leq \log(3) + 2\mathbf{E}\left[\log C(A, \boldsymbol{b}, \boldsymbol{c})\right] \\
&\leq O(\log(mn/\sigma)).
\end{aligned}
$$

7. Geometric Analysis of δ

To prove Lemma 6.1, we use the following lemma which says that if the value of $\boldsymbol{c}\boldsymbol{x}$ is close to optimal, then $\boldsymbol{x}$ must be close to $\boldsymbol{x}^*$.

Lemma 7.1. *For a linear program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$ with unique optimal primal and dual solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$, let*

$$\xi(A, \boldsymbol{b}, \boldsymbol{c}) = \min\left(\alpha_P(A, \boldsymbol{b}, \boldsymbol{c}), \alpha_D(A, \boldsymbol{b}, \boldsymbol{c}), \beta_P(A, \boldsymbol{b}, \boldsymbol{c}), \beta_D(A, \boldsymbol{b}, \boldsymbol{c})\right).$$

Then,

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_\infty \leq \boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x})\left(\frac{1 + \|A\|}{\xi(A, \boldsymbol{b}, \boldsymbol{c})\min(\gamma(A, \boldsymbol{b}, \boldsymbol{c}), 1)}\right)$$

Proof (of Lemma 6.1). Assuming

$$c(\boldsymbol{x}^* - \boldsymbol{x}) < \frac{\xi(A, \boldsymbol{b}, \boldsymbol{c})^2 \gamma(A, \boldsymbol{b}, \boldsymbol{c})}{2 \max(1, \sqrt{n} \, \|A\|) \, (1 + \|A\|)},$$

we need to show $B(\boldsymbol{x}) = B$ and $N(\boldsymbol{x}) = N$. From Lemma 7.1, we have

$$\begin{aligned}
\|\boldsymbol{x}^* - \boldsymbol{x}\|_\infty &< \frac{\xi^2 \gamma}{2 \max(1, \sqrt{n} \, \|A\|) \, (1 + \|A\|)} \left(\frac{1 + \|A\|}{\xi \min(\gamma, 1)} \right) \\
&\leq \frac{\xi}{2 \max(1, \sqrt{n} \, \|A\|)} \leq \frac{\xi}{2}.
\end{aligned} \tag{5}$$

We then have

(a) for $i \in B$, $x_i > x_i^* - \xi/2 > \alpha_P(A, \boldsymbol{b}, \boldsymbol{c}) - \xi/2 > \xi/2$,
(b) for $i \notin B$, $x_i < \xi/2$,

As $|A_{j,:}(\boldsymbol{x}^* - \boldsymbol{x})| \leq \|A\|_\infty \|\boldsymbol{x}^* - \boldsymbol{x}\|_\infty \leq \sqrt{n} \, \|A\| \, \|\boldsymbol{x}^* - \boldsymbol{x}\|_\infty \leq \xi/2$, by inequality (5), we also have

(c) for $j \in N$, $b_j - A_{j,:}\boldsymbol{x} < \xi/2$,
(d) for $j \notin N$, $b_j - A_{j,:}\boldsymbol{x} = b_j - A_{j,:}\boldsymbol{x}^* - A_{j,:}(\boldsymbol{x} - \boldsymbol{x}^*) > \beta_P(A, \boldsymbol{b}, \boldsymbol{c}) - \xi/2 > \xi/2$.

So, the smallest n values in $\{\boldsymbol{x}_i\} \cup \{\boldsymbol{b}_j - A_{j,:}\boldsymbol{x}\}$ are those indexed by $\bar{B}$ and N.

The proof of Lemma 7.1 relies on the following technical lemmas.

Lemma 7.2. *For $\boldsymbol{x}$ a feasible point of a bounded linear program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$, with unique optimal primal and dual solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$,*

$$c(\boldsymbol{x}^* - \boldsymbol{x}) \geq \alpha_D(A, \boldsymbol{b}, \boldsymbol{c}) \, \|A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x})\|.$$

Proof. As $\boldsymbol{y}^* A \geq \boldsymbol{c}$, we have

$$\begin{aligned}
c(\boldsymbol{x}^* - \boldsymbol{x}) &\geq c\boldsymbol{x}^* - \boldsymbol{y}^* A \boldsymbol{x} \\
&= \boldsymbol{y}^* A \boldsymbol{x}^* - \boldsymbol{y}^* A \boldsymbol{x} \qquad \text{(by strong duality)} \\
&= \boldsymbol{y}^* A (\boldsymbol{x}^* - \boldsymbol{x}) \\
&= \boldsymbol{y}_N^* A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x}),
\end{aligned}$$

as $\boldsymbol{y}^*$ is zero outside of N. As $A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x})$ is non-negative, we may conclude that

$$\boldsymbol{y}_N^* A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x}) \geq \alpha_D(A, \boldsymbol{b}, \boldsymbol{c}) \, \|A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x})\|_1 \geq \alpha_D(A, \boldsymbol{b}, \boldsymbol{c}) \, \|A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x})\|.$$

Lemma 7.3. *For $\boldsymbol{x}$ a feasible point of a bounded linear program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$, with unique optimal primal and dual solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$,*

$$\|A_{N,B}(\boldsymbol{x}_B^* - \boldsymbol{x}_B)\| \geq \gamma(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_B^* - \boldsymbol{x}_B\|_\infty .$$

Proof. For any $k \in B$, let $\boldsymbol{q}$ be the unit row vector for which $\boldsymbol{q}A_{N,B-k} = \boldsymbol{0}$. Then,

$$\boldsymbol{q}A_{N,B}\boldsymbol{x} = \boldsymbol{q}\left(A_{N,k}x_k + A_{N,B-k}\boldsymbol{x}_{B-k}\right) = x_k\boldsymbol{q}A_{N,k} = x_k\mathbf{dist}\left(A_{N,k}, \mathbf{Span}\left(A_{N,B-k}\right)\right).$$

So,

$$\begin{aligned}
\|A_{N,B}(\boldsymbol{x}_B^* - \boldsymbol{x}_B)\| &\geq |\boldsymbol{q}A_{N,B}(\boldsymbol{x}_B^* - \boldsymbol{x}_B)| \\
&= |x_k^* - x_k|\,\mathbf{dist}\left(A_{N,k}, \mathbf{Span}\left(A_{N,B-k}\right)\right) \\
&\geq |x_k^* - x_k|\,\gamma(A, \boldsymbol{b}, \boldsymbol{c}).
\end{aligned}$$

Lemma 7.4. *For $\boldsymbol{x}$ a feasible point of a bounded linear program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$, with unique optimal primal and dual solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$,*

$$\frac{\boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x})}{\beta_D(A, \boldsymbol{b}, \boldsymbol{c})} \geq \|\boldsymbol{x}_{\bar{B}}\| .$$

Proof. As $\boldsymbol{y}^*A_{:,B} = \boldsymbol{c}_B$, $(\boldsymbol{y}^*A - \boldsymbol{c})\boldsymbol{x} = (\boldsymbol{y}^*A_{:,\bar{B}} - \boldsymbol{c}_{\bar{B}})\boldsymbol{x}_{\bar{B}}$.

As every entry in $\boldsymbol{y}^*A_{:,\bar{B}} - \boldsymbol{c}_{\bar{B}}$ is at least $\beta_D(A, \boldsymbol{b}, \boldsymbol{c})$ and $\boldsymbol{x} \geq 0$, we have

$$(\boldsymbol{y}^*A - \boldsymbol{c})\boldsymbol{x} = (\boldsymbol{y}^*A_{:,\bar{B}} - \boldsymbol{c}_{\bar{B}})\boldsymbol{x}_{\bar{B}} \geq \beta_D(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_{\bar{B}}\|_1 \geq \beta_D(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_{\bar{B}}\| . \quad (6)$$

As $\boldsymbol{y}^* \geq 0$,

$$\boldsymbol{c}\left(\boldsymbol{x}^* - \boldsymbol{x}\right) = \boldsymbol{y}^*\boldsymbol{b} - \boldsymbol{c}\boldsymbol{x} \geq \boldsymbol{y}^*A\boldsymbol{x} - \boldsymbol{c}\boldsymbol{x} \geq \beta_D(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_{\bar{B}}\| ,$$

where the last inequality follows from (6).

Proof (of Lemma 7.1). Applying the triangle inequality, we observe

$$\|A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x})\| \geq \|A_{N,B}(\boldsymbol{x}_B^* - \boldsymbol{x}_B)\| - \|A_{N,\bar{B}}(\boldsymbol{x}_{\bar{B}}^* - \boldsymbol{x}_{\bar{B}})\| .$$

We can bound the first of these terms by applying Lemma 7.3, and the second by observing $\boldsymbol{x}_{\bar{B}}^* = 0$ and $\|A_{N,\bar{B}}\boldsymbol{x}_{\bar{B}}\| \leq \|A_{N,\bar{B}}\| \|\boldsymbol{x}_{\bar{B}}\|$, thereby proving

$$\|A_{N,:}(\boldsymbol{x}^* - \boldsymbol{x})\| \geq \gamma(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_B^* - \boldsymbol{x}_B\|_\infty - \|A_{N,\bar{B}}\| \|\boldsymbol{x}_{\bar{B}}\| .$$

Now applying Lemma 7.2, we obtain

$$\boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x}) \geq \alpha_D(A, \boldsymbol{b}, \boldsymbol{c})\left(\gamma(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_B^* - \boldsymbol{x}_B\|_\infty - \|A_{N,\bar{B}}\| \|\boldsymbol{x}_{\bar{B}}\|\right),$$

which implies

$$\alpha_D(A, \boldsymbol{b}, \boldsymbol{c}) \|A_{N,\bar{B}}\| \|\boldsymbol{x}_{\bar{B}}\| + \boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x}) \geq \alpha_D(A, \boldsymbol{b}, \boldsymbol{c})\gamma(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_B^* - \boldsymbol{x}_B\|_\infty .$$

As Lemma 7.4 implies $\boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x})/\beta_D(A, \boldsymbol{b}, \boldsymbol{c}) \geq \|\boldsymbol{x}_{\bar{B}}\|$ and $\|A\| \geq \|A_{N,\bar{B}}\|$, we obtain

$$\left(1 + \frac{\alpha_D(A, \boldsymbol{b}, \boldsymbol{c})}{\beta_D(A, \boldsymbol{b}, \boldsymbol{c})} \|A\| \right) \boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x}) \geq \alpha_D(A, \boldsymbol{b}, \boldsymbol{c}) \gamma(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_B^* - \boldsymbol{x}_B\|_\infty,$$

which implies

$$\frac{\boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x})\left(1 + \|A\| \right)}{\min\left(\alpha_D(A, \boldsymbol{b}, \boldsymbol{c}), \beta_D(A, \boldsymbol{b}, \boldsymbol{c}) \right)} \geq \gamma(A, \boldsymbol{b}, \boldsymbol{c}) \|\boldsymbol{x}_B^* - \boldsymbol{x}_B\|_\infty.$$

As $\|\boldsymbol{x}^* - \boldsymbol{x}\|_\infty = \max\left(\|\boldsymbol{x}_B^* - \boldsymbol{x}_B\|_\infty, \|\boldsymbol{x}_{\bar{B}}^* - \boldsymbol{x}_{\bar{B}}\|_\infty \right)$, the lemma now follows from this last inequality and Lemma 7.4, which implies

$$\frac{\boldsymbol{c}(\boldsymbol{x}^* - \boldsymbol{x})}{\beta_D(A, \boldsymbol{b}, \boldsymbol{c})} \geq \|\boldsymbol{x}_{\bar{B}}\| \geq \|\boldsymbol{x}_{\bar{B}}\|_\infty = \|\boldsymbol{x}_{\bar{B}}^* - \boldsymbol{x}_{\bar{B}}\|_\infty.$$

8. Bounds on α, β, and γ

For this section, we let μ_A, $\mu_{\boldsymbol{b}}$ and $\mu_{\boldsymbol{c}}$ denote the densities according to which A, $\boldsymbol{b}$ and $\boldsymbol{c}$ are distributed under the conditions of Theorem 5.2. For an index j or set of indices N, we let μ_{b_j} and $\mu_{\boldsymbol{b}_N}$ denote the induced distributions on b_j and $\boldsymbol{b}_N$, and we extend this notational convention to sub-matrices of A and sub-vectors of $\boldsymbol{c}$.

The idea behind our proofs of Lemma 6.2, 6.3 and 6.4 is that for any configuration of A, $\boldsymbol{b}$ and $\boldsymbol{c}$ in which α, β, or γ is small, there are many nearby configurations in which the term is not too small. As Gaussian densities do not fall off too quickly, this nearby configuration will have approximately the same probability as the original. To make this idea rigorous, we establish mappings pairing configurations in which these terms are small with configurations in which these terms are not. We then use these mappings to show that the Gaussian probability of the configurations in which the terms are not small is much larger than those in which they are.

To show that it is unlikely that β_P is small, we hold A, $\boldsymbol{y}^*$, $\boldsymbol{c}$, and $\boldsymbol{x}^*$ constant, and map those $\boldsymbol{b}_j$'s that are close to $A_{j,:}\boldsymbol{x}^*$ to be a little further away. To show that it is unlikely that α_P is small, we hold A, $\boldsymbol{y}^*$ and $\boldsymbol{c}$ constant, and map small non-zero entries of $\boldsymbol{x}^*$ to larger values while simultaneously mapping the entries of $\boldsymbol{b}$ so as to preserve the tight constraints and maintain slack in the others. To show that it is unlikely that γ is small, we hold $\boldsymbol{x}^*$, $\boldsymbol{y}^*$, and the slack components of $\boldsymbol{b}$ and $\boldsymbol{c}$ constant. We then vary $A_{N,B}$ slightly, changing $\boldsymbol{b}_N$ and $\boldsymbol{c}_B$ accordingly. As each slight motion described only induces a slight motion in the other components, we can prove that each configuration obtained has similar probability.

To turn these intuitive arguments into proofs, we need four tools:

1. a bound on the smoothness of the Gaussian density,
2. a bound on the probability that a random variable is small given that its density is smooth near zero,
3. a lemma making rigorous the change of variables implicitly used in the intuitive arguments, and
4. a proof that the probability of an event can be bounded by the maximum of its probability over the sets in a partition of its probability space.

Each of these tools is relatively simple, and the last should be obvious for finite partitions. The bound on the smoothed complexity of the simplex method [ST01] uses each of these tools along with some others. It is our hope that the reader would have an easier time understanding the proofs in [ST01] after having read this section.

We now develop these four tools, and at the end of the section apply them to the proofs of the bounds on α, β and γ.

We make use of the following elementary bound on the smoothness of Gaussians:

Lemma 8.1 (Smoothness of Gaussians). *Let $\mu(\boldsymbol{x})$ be a Gaussian distribution in $\mathbb{R}^n$ of variance σ^2 centered at a point of norm at most 1. If $\mathbf{dist}\,(\boldsymbol{x}, \boldsymbol{y}) < \epsilon \leq 1$, then*

$$\frac{\mu(\boldsymbol{y})}{\mu(\boldsymbol{x})} \geq e^{-\frac{\epsilon(\|x\|+2)}{\sigma^2}}.$$

Proof. Let $\hat{\boldsymbol{x}}$ be the center of the distribution. We compute

$$
\begin{aligned}
\frac{\mu(\boldsymbol{y})}{\mu(\boldsymbol{x})} &= e^{\frac{-1}{2\sigma^2}\left(\|\boldsymbol{y}-\hat{\boldsymbol{x}}\|^2-\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2\right)} \\
&\geq e^{\frac{-1}{2\sigma^2}\left((\|\boldsymbol{y}-\boldsymbol{x}\|+\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|)^2-\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2\right)} &&\text{by the triangle inequality} \\
&= e^{\frac{-1}{2\sigma^2}\left(2\|\boldsymbol{y}-\boldsymbol{x}\|\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|+\|\boldsymbol{y}-\boldsymbol{x}\|^2\right)} \\
&\geq e^{\frac{-1}{2\sigma^2}\left(2\epsilon\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|+\epsilon^2\right)} \\
&\geq e^{\frac{-1}{2\sigma^2}\left(2\epsilon(\|\boldsymbol{x}\|+1)+\epsilon^2\right)} &&\text{as } \|\hat{\boldsymbol{x}}\| \leq 1 \\
&\geq e^{\frac{-1}{2\sigma^2}\left(2\epsilon(\|\boldsymbol{x}\|+1)+\epsilon\right)} &&\text{as } \epsilon \leq 1 \\
&\geq e^{-\frac{\epsilon(\|x\|+2)}{\sigma^2}}.
\end{aligned}
$$

We remark that this lemma is the only fact about Gaussian random variables used in this section. Thus, one could obtain results of a similar character for any distribution that satisfies properties similar to those derived for Gaussian random vectors above, provided that the vectors are comprised of independently distributed entries.

The argument by which we obtain probability bounds from comparing configurations is encapsulated in the following lemma, which is used in each of the three

proofs. This lemma essentially says that if the distribution of a random variable is relatively flat near a point, then the variable is unlikely to lie too close to that point.

Lemma 8.2 (Smooth distributions unlikely small). *Let x be a real random variable distributed according to a density ρ for which there exist constants α and c such that*

$$0 \le x \le x' \le \alpha \ \text{implies} \quad \frac{\rho(x')}{\rho(x)} \ge c.$$

Then, for $\epsilon < \alpha$,

$$\Pr\left[x \in [0, \epsilon] \big| x \in [0, \alpha]\right] \le \frac{\epsilon}{c\alpha}.$$

In particular,

$$\Pr\left[x \in [0, \epsilon]\right] \le \frac{\epsilon}{c\alpha}.$$

Proof. From the definition of conditional probability, we have

$$\Pr\left[x \in [0, \epsilon] \big| x \in [0, \alpha]\right] = \frac{\int_0^\epsilon \rho(x)\, dx}{\int_0^\alpha \rho(x)\, dx}.$$

Applying the change of variables $y = (\epsilon/\alpha)x$, we compute

$$\int_0^\alpha \rho(x)\, dx = (\alpha/\epsilon) \int_0^\epsilon \rho((\alpha/\epsilon)y)\, dy \ge (\alpha/\epsilon) \int_0^\epsilon c\rho(y)\, dy.$$

From which the lemma follows.

For example, we can use the previous lemma to derive a bound on the probability that a Gaussian random variable is greater than $t + \epsilon$ given that it is greater than t:

Lemma 8.3 (Comparison of Gaussian tails). *Let x be a Gaussian random variable of variance $\sigma^2 \le 1$ and mean of absolute value at most 1. For $\epsilon \ge 0$, $\tau \ge 1$ and $t \le \tau$,*

$$\Pr\left[x \le t + \epsilon \big| x \ge t\right] \le \frac{\epsilon\tau}{\sigma^2} e^{\frac{\epsilon(\tau+3)}{\sigma^2}} \quad \text{and} \quad \Pr\left[x \ge t + \epsilon \big| x \ge t\right] \ge 1 - \frac{\epsilon\tau}{\sigma^2} e^{\frac{\epsilon(\tau+3)}{\sigma^2}}.$$

Proof. It suffices to prove the first bound. Let μ be the density function of x. Let $\alpha = \sigma^2/\tau$, and note $\alpha \le 1$. For $\epsilon \ge \alpha$ the lemma is vacuous. For $\epsilon < \alpha$, we will show that

$$t \le x < x' \le t + \alpha \ \text{implies} \ \frac{\mu(x')}{\mu(x)} \ge e^{-\frac{\epsilon(\tau+3)}{\sigma^2}}, \tag{7}$$

and then apply Lemma 8.2 to finish the proof. For $t < -1$, (7) is trivial as μ is monotone increasing on $[t, t+\alpha]$. For $t \ge -1$, we have $\|x\| \le \tau + 1$; so, (7) follows from Lemma 8.1.

Finally, we note that our intuitive explanation of the proofs of Lemmas 6.2, 6.3 and 6.4 implicitly used a change of variables: instead of reasoning in terms of the variables A, $\boldsymbol{b}$ and $\boldsymbol{c}$, we found it more convenient to think of $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ as quantities to fix or vary. We now introduce the machinery that enables us to reason in terms of these variables. We begin by observing that for any sets B and N, not necessarily the combinatorial type of $(A, \boldsymbol{b}, \boldsymbol{c})$, we can introduce variables $\boldsymbol{x}_B^*$ and $\boldsymbol{y}_N^*$, not necessarily the optimal primal and dual solutions, and define

$$\boldsymbol{b}_N = A_{N,B} \boldsymbol{x}_B^* \quad \text{and} \quad \boldsymbol{c}_B = \boldsymbol{y}_N^* A_{N,B}.$$

We can then compute probabilities in these new variables by observing that the joint density of A, $\boldsymbol{x}_B^*$, $\boldsymbol{y}_N^*$, $\boldsymbol{b}_{\bar{N}}$, and $\boldsymbol{c}_{\bar{B}}$ is

$$\mu_A(A)\mu_{b_N}(A_{N,B}\boldsymbol{x}_B^*)\mu_{c_B}(\boldsymbol{y}_N^* A_{N,B})\mu_{b_{\bar{N}}}(\boldsymbol{b}_{\bar{N}})\mu_{c_{\bar{B}}}(\boldsymbol{c}_{\bar{B}})\mathbf{det}\,(A_{N,B})^2\,.$$

To see why this is true, recall that probabilities are best understood as integrals, and that the probability of an event $\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})$ is

$$\int_{A,\boldsymbol{b},\boldsymbol{c}} [\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})]\, \mu_A(A)\mu_b(\boldsymbol{b})\mu_c(\boldsymbol{c})\, dA\, d\boldsymbol{b}\, d\boldsymbol{c} \tag{8}$$

To express this integral in the new variables, we first compute the Jacobian of the change of variables, which is

$$\left| \mathbf{det}\left(\frac{\partial(A, \boldsymbol{b}_{\bar{N}}, \boldsymbol{c}_{\bar{B}}, \boldsymbol{b}_N, \boldsymbol{c}_B)}{\partial(A, \boldsymbol{b}_{\bar{N}}, \boldsymbol{c}_{\bar{B}}, \boldsymbol{x}_B^*, \boldsymbol{y}_N^*)} \right) \right| = \mathbf{det}\,(A_{N,B})^2\,;$$

so,

$$dA\, d\boldsymbol{b}\, d\boldsymbol{c} = \left| \mathbf{det}\left(\frac{\partial(A, \boldsymbol{b}_{\bar{N}}, \boldsymbol{c}_{\bar{B}}, \boldsymbol{b}_N, \boldsymbol{c}_B)}{\partial(A, \boldsymbol{b}_{\bar{N}}, \boldsymbol{c}_{\bar{B}}, \boldsymbol{x}_B^*, \boldsymbol{y}_N^*)} \right) \right| dA\, d\boldsymbol{b}_{\bar{N}}\, d\boldsymbol{c}_{\bar{B}}\, d\boldsymbol{x}_B^*\, d\boldsymbol{y}_N^*$$
$$= \mathbf{det}\,(A_{N,B})^2\, dA\, d\boldsymbol{b}_{\bar{N}}\, d\boldsymbol{c}_{\bar{B}}\, d\boldsymbol{x}_B^*\, d\boldsymbol{y}_N^*\,,$$

and

$$(8) = \int_{A,\boldsymbol{b}_{\bar{N}},\boldsymbol{c}_{\bar{B}},\boldsymbol{x}_B^*,\boldsymbol{y}_N^*} [\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})]$$
$$\cdot \mu_A(A)\mu_{b_N}(A_{N,B}\boldsymbol{x}_B^*)\mu_{c_B}(\boldsymbol{y}_N^* A_{N,B})\mu_{b_{\bar{N}}}(\boldsymbol{b}_{\bar{N}})\mu_{c_{\bar{B}}}(\boldsymbol{c}_{\bar{B}})\mathbf{det}\,(A_{N,B})^2$$
$$\cdot dA\, d\boldsymbol{b}_{\bar{N}}\, d\boldsymbol{c}_{\bar{B}}\, d\boldsymbol{x}_B^*\, d\boldsymbol{y}_N^*\,.$$

While we can define this change of variables for any sets B and N, we will of course only apply this change of variables to programs of type (B, N). If we let $\mathrm{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})$ denote the set of $(A, \boldsymbol{b}, \boldsymbol{c})$ of type (B, N), then we can express the probability of $[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})$ and $\mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})]$ as

$$\int_{A,\boldsymbol{b},\boldsymbol{c}:\mathcal{F}(A,\boldsymbol{b},\boldsymbol{c})} [\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})]\, \mu_A(A)\mu_b(\boldsymbol{b})\mu_c(\boldsymbol{c})\, dA\, d\boldsymbol{b}\, d\boldsymbol{c}$$
$$= \sum_{B,N} \int_{A,\boldsymbol{b},\boldsymbol{c}:\mathrm{Type}_{B,N}(A,\boldsymbol{b},\boldsymbol{c})} [\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})]\, \mu_A(A)\mu_b(\boldsymbol{b})\mu_c(\boldsymbol{c})\, dA\, d\boldsymbol{b}\, d\boldsymbol{c}\,,$$

and then apply the change of variables corresponding to (B, N) to evaluate the integral over $\text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})$ on the right. In fact, in each of our proofs, we will actually bound

$$\max_{B,N} \Pr\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})|\text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})\right].$$

To see that this upper bounds the probability of $[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})$ and $\mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})]$, we prove

Claim.

$$\Pr\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c}) \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right] \leq \max_{B,N} \Pr\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})|\text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})\right].$$

Proof.

$$\Pr\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c}) \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$
$$= \sum_{B,N} \Pr\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c}) \text{ and } \text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$
$$= \sum_{B,N} \Pr\left[\text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})\right] \Pr\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})|\text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$
$$\leq \max_{B,N} \Pr\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})|\text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})\right],$$

as $\sum_{B,N} \Pr\left[\text{Type}_{B,N}(A, \boldsymbol{b}, \boldsymbol{c})\right] \leq 1$.

We summarize this discussion in the following lemma:

Lemma 8.4 (Change of variables). *Let $\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})$ be an event. Then,*

$$\Pr_{A,\boldsymbol{b},\boldsymbol{c}}\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c}) \text{ and } \mathcal{F}(A, \boldsymbol{b}, \boldsymbol{c})\right]$$
$$\leq \max_{B,N} \Pr_{A,\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{b}_{\bar{N}},\boldsymbol{c}_{\bar{B}}}\left[\mathcal{E}(A, \boldsymbol{b}, \boldsymbol{c})|A_{\bar{N},:}\boldsymbol{x}^* \leq \boldsymbol{b}_{\bar{N}} \text{ and } \boldsymbol{y}^*A_{:,\bar{B}} \geq \boldsymbol{c}_{\bar{B}}\right],$$

where A, $\boldsymbol{x}^*$, $\boldsymbol{y}^*$, $\boldsymbol{b}_{\bar{N}}$ and $\boldsymbol{c}_{\bar{B}}$ have joint density

$$\mu_A(A)\mu_{b_N}(A_{N,B}\boldsymbol{x}_B^*)\mu_{c_B}(\boldsymbol{y}_N^*A_{N,B})\mu_{b_{\bar{N}}}(\boldsymbol{b}_{\bar{N}})\mu_{c_{\bar{B}}}(\boldsymbol{c}_{\bar{B}})\boldsymbol{det}\left(A_{N,B}\right)^2.$$

All of our proofs begin by fixing some subset of the variables, and then proving a probability bound for any configuration of the fixed variables. This amounts to proving a probability upper bound by dividing the probability space into a number of regions, and proving that the bound holds in each of these regions. Formally, we are using the fact:

Proposition 8.1 (Upper bound by max of probabilities). *Let X and Y be random variables distributed according to an integrable density function $\mu(X,Y)$ and let $\mathcal{E}(X,Y)$ be an event. Then*

$$\Pr_{X,Y}\left[\mathcal{E}(X,Y)\right] \leq \max_{y} \Pr_{X,Y}\left[\mathcal{E}(X,Y)|Y=y\right] \stackrel{\text{def}}{=} \max_{Y}\Pr_{X}\left[\mathcal{E}(X,Y)|Y\right].$$

Proof. By Tonelli's Theorem, we have

$$\Pr_{X,Y}\left[\mathcal{E}(X,Y)\right] = \int_{X,Y}\left[\mathcal{E}(X,Y)\right]\mu(X,Y)\,dX\,dY$$

$$= \int_{Y}\left(\int_{X}\left[\mathcal{E}(X,Y)\right]\mu(X,Y)\,dX\right)dY$$

$$= \int_{Y}\left(\int_{X}\mu(X,Y)\,dX\right)\left(\frac{\int_{X}\left[\mathcal{E}(X,Y)\right]\mu(X,Y)\,dX}{\left(\int_{X}\mu(X,Y)\,dX\right)}\right)dY$$

$$= \int_{Y}\left(\int_{X}\mu(X,Y)\,dX\right)\left(\Pr_{X}\left[\mathcal{E}(X,Y)|Y\right]\right)dY$$

$$\leq \max_{Y}\Pr_{X}\left[\mathcal{E}(X,Y)|Y\right],$$

as

$$\int_{Y}\left(\int_{X}\mu(X,Y)\,dX\right)dY = 1.$$

Having established these tools, we now proceed with the proofs of Lemmas 6.2, 6.3 and 6.4.

Proof (of Lemma 6.3 (Probability of small β)). By Lemma 8.4, it suffices to bound

$$\max_{B,N}\Pr_{A,\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{b}_{\bar{N}},\boldsymbol{c}_{\bar{B}}}\left[\begin{array}{l}\beta_P(A,\boldsymbol{b},\boldsymbol{c}) \leq \frac{\epsilon}{\max(1,\|A\|\|\boldsymbol{x}^*\|)} \\ \left| A_{\bar{N},:}\boldsymbol{x}^* \leq \boldsymbol{b}_{\bar{N}} \text{ and } \boldsymbol{y}^* A_{:,\bar{B}} \geq \boldsymbol{c}_{\bar{B}} \right.\end{array}\right].$$

By Proposition 8.1, it suffices to prove that for all B, N, A, $\boldsymbol{x}_B^*$, $\boldsymbol{y}_N^*$ and $\boldsymbol{c}_{\bar{B}}$ satisfying $\boldsymbol{y}^* A_{:,\bar{B}} \geq \boldsymbol{c}_{\bar{B}}$,

$$\Pr_{\boldsymbol{b}_{\bar{N}}}\left[\exists j \in \bar{N} : b_j - A_{j,:}\boldsymbol{x}^* \leq \epsilon' \mid \forall j : b_j - A_{j,:}\boldsymbol{x}^* \geq 0\right]$$

$$\leq \sum_{j \in \bar{N}}\Pr_{b_j}\left[b_j - A_{j,:}\boldsymbol{x}^* \leq \epsilon' \mid \forall j : b_j - A_{j,:}\boldsymbol{x}^* \geq 0\right]$$

$$= \sum_{j \in \bar{N}}\Pr_{b_j}\left[b_j - A_{j,:}\boldsymbol{x}^* \leq \epsilon' \mid b_j - A_{j,:}\boldsymbol{x}^* \geq 0\right]$$

$$= \sum_{j \in \bar{N}}\Pr_{b_j}\left[b_j \leq A_{j,:}\boldsymbol{x}^* + \epsilon' \mid b_j \geq A_{j,:}\boldsymbol{x}^*\right]$$

$$\leq \frac{m\epsilon'\,|A_{j,:}\boldsymbol{x}^*|}{\sigma^2}e^{\frac{\epsilon'(|A_{j,:}\boldsymbol{x}^*|+3)}{\sigma^2}} \qquad\qquad \text{by Lemma 8.3.}$$

$$\leq \frac{m\epsilon'\,(\|A\|\,\|\boldsymbol{x}^*\|)}{\sigma^2}e^{\frac{\epsilon'(\|A\|\|\boldsymbol{x}^*\|+3)}{\sigma^2}}.$$

Setting $\epsilon' = \epsilon/\max(1, \|A\| \|\boldsymbol{x}^*\|)$, and observing that the lemma is vacuously true for $\epsilon > \sigma^2/4m$, we deduce

$$\Pr_{\boldsymbol{b}_{\bar{N}}}\left[\exists j \in \bar{N} : b_j - A_{j,:}\boldsymbol{x}^* \le \epsilon \;\;\middle|\;\; \forall j : b_j - A_{j,:}\boldsymbol{x}^* \ge 0\right] \le \frac{m\epsilon}{\sigma^2}e^{\frac{4\epsilon}{\sigma^2}} \le \frac{em\epsilon}{\sigma^2} \le \frac{4m\epsilon}{\sigma^2},$$

for $\epsilon < \sigma^2/4m$.

Proof (of Lemma 6.4 (Probability of small γ)). By Lemma 8.4, it suffices to bound

$$\max_{B,N} \Pr_{A, \boldsymbol{x}_B^*, \boldsymbol{y}_N^*, \boldsymbol{b}_{\bar{N}}, \boldsymbol{c}_{\bar{B}}} \left[\begin{array}{l} \gamma(A, \boldsymbol{b}, \boldsymbol{c}) \le \frac{\epsilon}{\left(1+\|\boldsymbol{x}^*\|^2 + \|\boldsymbol{y}^*\|^2\right)(\|A\|+3)} \\[2mm] \left| A_{\bar{N},:}\boldsymbol{x}^* \le \boldsymbol{b}_{\bar{N}} \text{ and } \boldsymbol{y}^*A_{:,\bar{B}} \ge \boldsymbol{c}_{\bar{B}} \right. \end{array} \right].$$

By Proposition 8.1, it suffices to prove that for all B, N, $A_{\overline{N,B}}$, $\boldsymbol{b}_{\bar{N}}$, $\boldsymbol{c}_{\bar{B}}$, $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ satisfying $A_{\bar{N},B}\boldsymbol{x}_B^* \le \boldsymbol{b}_{\bar{N}}$ and $\boldsymbol{y}_N^*A_{N,\bar{B}} \ge \boldsymbol{c}_{\bar{B}}$,

$$\Pr_{A_{N,B}}\left[\gamma(A, \boldsymbol{b}, \boldsymbol{c}) \le \epsilon\right] \le \frac{3e\epsilon n\left(1+\|\boldsymbol{x}^*\|^2 + \|\boldsymbol{y}^*\|^2\right)(\|A_{N,B}\|+3)}{\sigma^2}, \qquad (9)$$

where we note that having fixed $A_{\overline{N,B}}$, $\boldsymbol{b}_{\bar{N}}$, $\boldsymbol{c}_{\bar{B}}$, $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$, the induced distribution on $A_{N,B}$ is

$$\mu_{A_{N,B}}(A_{N,B})\mu_{b_N}(A_{N,B}\boldsymbol{x}_B^*)\mu_{c_B}(\boldsymbol{y}_N^*A_{N,B})\det(A_{N,B})^2.$$

To prove (9), we show that for all $k \in B$ and all $A_{N,B-k}$,

$$\Pr_{A_{N,k}}\left[\mathbf{dist}\left(A_{N,k}, \mathbf{Span}\left(A_{N,B-k}\right)\right) \le \epsilon\right] \le \frac{3e\epsilon\left(1+\|\boldsymbol{x}^*\|^2 + \|\boldsymbol{y}^*\|^2\right)(\|A_{N,B}\|+3)}{\sigma^2},$$
(10)

and apply a union bound over $k \in B$. Having fixed $A_{N,B-k}$, we may express $A_{N,k}$ as $\boldsymbol{a} + t\boldsymbol{q}$ where $\boldsymbol{a} \in \mathbf{Span}\left(A_{N,B-k}\right)$, $\boldsymbol{q}$ is the unit vector orthogonal to $\mathbf{Span}\left(A_{N,B-k}\right)$ and $t \in \mathbb{R}$. With this representation, we have $|t| = \mathbf{dist}\left(A_{N,k}, \mathbf{Span}\left(A_{N,B-k}\right)\right)$ and $\det(A_{N,B}) = ct$, where c is some constant depending only on $A_{N,B-k}$.

By the symmetry of $\boldsymbol{q}$ with $-\boldsymbol{q}$, we can prove (10) by bounding the probability that t is less than ϵ given that t is at least 0. Thus, we prove (10) by observing $\|A_{N,B}\| \ge \|A_{N,B-k}, \boldsymbol{a}\|$ and showing

$$\max_{\boldsymbol{a} \in \mathbf{Span}(A_{N,B-k})} \Pr_t\left[t \le \epsilon \middle| t \ge 0\right] \le \frac{3e\epsilon\left(1+\|\boldsymbol{x}^*\|^2 + \|\boldsymbol{y}^*\|^2\right)(\|A_{N,B-k}, \boldsymbol{a}\|+3)}{\sigma^2}$$
(11)

where the induced distribution on t is proportional to

$$\rho(t) \stackrel{\text{def}}{=} \mu_{A_{N,k}}(\boldsymbol{a} + t\boldsymbol{q})\mu_{b_N}(A_{N,B-k}\boldsymbol{x}_{B-k}^* + (\boldsymbol{a} + t\boldsymbol{q})x_k^*)\mu_{c_k}\left(\boldsymbol{y}_N^*(\boldsymbol{a} + t\boldsymbol{q})\right)t^2. \quad (12)$$

We now set

$$\alpha = \frac{\sigma^2}{3\left(1 + \|\boldsymbol{x}^*\|^2 + \|\boldsymbol{y}^*\|^2\right)\left(\|A_{N,B-k}, \boldsymbol{a}\| + 3\right)}$$

and prove

$$0 \le t \le t' \le \alpha \text{ implies } \frac{\rho(t')}{\rho(t)} \ge 1/e, \tag{13}$$

from which (11) follows by Lemma 8.2.

To prove (13), we observe

1. As $\mathbf{dist}\,(\boldsymbol{a} + t'\boldsymbol{q}, \boldsymbol{a} + t\boldsymbol{q}) \le t' - t \le \alpha \le 1$, we may apply Lemma 8.1 to show

$$\frac{\mu_{A_{N,k}}(\boldsymbol{a} + t'\boldsymbol{q})}{\mu_{A_{N,k}}(\boldsymbol{a} + t\boldsymbol{q})} \ge e^{\frac{-\alpha(\|\boldsymbol{a} + t\boldsymbol{q}\| + 2)}{\sigma^2}} \ge e^{\frac{-\alpha(\|\boldsymbol{a}\| + 3)}{\sigma^2}} \ge e^{-1/3}.$$

2. As

$$\left\|A_{N,B-k}\boldsymbol{x}_{B-k}^* + \boldsymbol{a}x_k^* + t\boldsymbol{q}x_k^*\right\| \le \left(\|A_{N,B-k}, \boldsymbol{a}\| + \alpha\right)\|\boldsymbol{x}_B^*\|,$$

and

$$\mathbf{dist}\,\left(A_{N,B-k}\boldsymbol{x}_{B-k}^* + \boldsymbol{a}x_k^* + t\boldsymbol{q}x_k^*, A_{N,B-k}\boldsymbol{x}_{B-k}^* + \boldsymbol{a}x_k^* + t'\boldsymbol{q}x_k^*\right) \\ = (t' - t)\,x_k^* \le \alpha x_k^*,$$

Lemma 8.1 implies

$$\frac{\mu_{b_N}(A_{N,B-k}\boldsymbol{x}_{B-k}^* + (\boldsymbol{a} + t'\boldsymbol{q})x_i^*)}{\mu_{b_N}(A_{N,B-k}\boldsymbol{x}_{B-k}^* + (\boldsymbol{a} + t\boldsymbol{q})x_i^*)} \ge e^{\frac{-\alpha x_k^*\left(\left(\|A_{N,B-k}, \boldsymbol{a}\| + \alpha\right)\|\boldsymbol{x}_B^*\| + 2\right)}{\sigma^2}} \ge e^{-1/3}.$$

3. As $|\boldsymbol{y}_N^*(\boldsymbol{a} + t\boldsymbol{q})| \le \|\boldsymbol{y}_N^*\|\,(\|\boldsymbol{a}\| + t\,\|\boldsymbol{q}\|) \le \|\boldsymbol{y}_N^*\|\,(\|\boldsymbol{a}\| + \alpha)$, and

$$\mathbf{dist}\,(\boldsymbol{y}_N^*(\boldsymbol{a} + t\boldsymbol{q}), \boldsymbol{y}_N^*(\boldsymbol{a} + t'\boldsymbol{q})) = (t' - t)\,|\boldsymbol{y}_N^*\boldsymbol{q}| \le \alpha\,\|\boldsymbol{y}_N^*\|,$$

Lemma 8.1 implies

$$\frac{\mu_{c_k}\left(\boldsymbol{y}_N^*(\boldsymbol{a} + t'\boldsymbol{q})\right)}{\mu_{c_k}\left(\boldsymbol{y}_N^*(\boldsymbol{a} + t\boldsymbol{q})\right)} \ge e^{\frac{-\alpha\|\boldsymbol{y}^*\|(\|\boldsymbol{y}^*\|(\|\boldsymbol{a}\| + \alpha) + 2)}{\sigma^2}} \ge e^{-1/3}.$$

Proof (of Lemma 6.2 (Probability of small α)). By Lemma 8.4, it suffices to bound

$$\max_{B,N}\,\Pr_{A,\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{b}_{\bar{N}},\boldsymbol{c}_{\bar{B}}}\left[\begin{array}{c}\alpha_P(A, \boldsymbol{b}, \boldsymbol{c}) \le \frac{\epsilon}{(\|A\| + 2)^2(\|\boldsymbol{x}_B^*\| + 1)} \\ \left| A_{\bar{N},:}\boldsymbol{x}^* \le \boldsymbol{b}_{\bar{N}} \text{ and } \boldsymbol{y}^*A_{:,\bar{B}} \ge \boldsymbol{c}_{\bar{B}}\end{array}\right].$$

By Proposition 8.1, it suffices to prove that for all B, N, A, $\boldsymbol{y}_N^*$ and $\boldsymbol{c}_{\bar{B}}$ satisfying $\boldsymbol{y}_N^*A_{N,\bar{B}} \ge \boldsymbol{c}_{\bar{B}}$,

$$\Pr_{\boldsymbol{x}_B^*, \boldsymbol{b}_{\bar{N}}}\left[\alpha(A, \boldsymbol{b}, \boldsymbol{c}) \le \frac{\epsilon}{(\|A\| + 2)^2(\|\boldsymbol{x}_B^*\| + 1)}\,\middle|\,A_{\bar{N},B}\boldsymbol{x}_B^* \le \boldsymbol{b}_{\bar{N}}\right] \le \frac{8\epsilon n(m + 1)}{\sigma^2}, \tag{14}$$

where we note that, fixing B, N, A, $\boldsymbol{y}^*$ and $\boldsymbol{c}_{\bar{B}}$ and conditioning upon $A_{\bar{N},B}\boldsymbol{x}_B^* \leq \boldsymbol{b}_{\bar{N}}$, the induced density on $\boldsymbol{x}_B^*$ is proportional to

$$\mu_{b_N}(A_{N,B}\boldsymbol{x}_B^*) \prod_{j \notin N} \Pr_{b_j}[b_j > A_{j,B}\boldsymbol{x}_B^*].$$

To prove (14) , we will show

$$\forall_{i \in B} \forall_{\boldsymbol{x}_{B-i}^*} \Pr_{x_i^*}\left[x_i^* \leq \epsilon \middle| A_{\bar{N},B}\boldsymbol{x}_B^* \leq \boldsymbol{b}_{\bar{N}}\right] \leq \frac{8\epsilon(m+1)\left(\|A\|+2\right)^2\left(\left\|\boldsymbol{x}_{B-i}^*\right\|+1\right)}{\sigma^2},$$
(15)

which implies

$$\max_{i \in B} \max_{\boldsymbol{x}_{B-i}^*} \Pr_{x_i^*}\left[x_i^* \leq \frac{\epsilon}{\left(\|A\|+2\right)^2\left(\left\|\boldsymbol{x}_{B-i}^*\right\|+1\right)} \middle| A_{\bar{N},B}\boldsymbol{x}_B^* \leq \boldsymbol{b}_{\bar{N}}\right] \leq \frac{8\epsilon(m+1)}{\sigma^2}.$$

We then observe $\frac{\epsilon}{(\|A\|+2)^2\left(\left\|\boldsymbol{x}_{B-i}^*\right\|+1\right)} \leq \frac{\epsilon}{(\|A\|+2)^2\left(\left\|\boldsymbol{x}_B^*\right\|+1\right)}$ and union bound over $i \in B$. To prove (15), we first note that having fixed $i \in B$ and $\boldsymbol{x}_{B-i}^*$, the induced density on x_i^* is proportional to

$$\rho(x_i^*) \stackrel{\text{def}}{=} \mu_{b_N}(A_{N,B-i}\boldsymbol{x}_{B-i}^* + A_{N,i}x_i^*) \prod_{j \notin N} \Pr_{b_j}\left[b_j > A_{j,B-i}\boldsymbol{x}_{B-i}^* + A_{j,i}x_i^*\right]$$

We now set

$$\alpha = \frac{\sigma^2}{4(m+1)\left(\|A\|+2\right)^2\left(\left\|\boldsymbol{x}_{B-i}^*\right\|+1\right)} \leq 1,$$

and prove that

$$0 \leq x_i \leq x_i' \leq \alpha \text{ implies } \frac{\rho(x_i')}{\rho(x_i)} \geq 1/2,$$
(16)

from which (15) follows by Lemma 8.2.

To prove (16), we note that for $0 \leq x_i \leq x_i' \leq \alpha$,

$$\mathbf{dist}\left(\left(A_{N,B-i}\boldsymbol{x}_{B-i}^* + A_{N,i}x_i\right), \left(A_{N,B-i}\boldsymbol{x}_{B-i}^* + A_{N,i}x_i'\right)\right)$$
$$= \|A_{N,i}\|(x_i' - x_i) \leq \|A\|\alpha \leq 1,$$

and

$$\left\|A_{N,B-i}\boldsymbol{x}_{B-i}^* + A_{N,i}x_i\right\| \leq \|A_{N,B-i}\|\left\|\boldsymbol{x}_{B-i}^*\right\| + \|A_{N,i}\|x_i \leq \|A\|\left(\left\|\boldsymbol{x}_{B-i}^*\right\|+1\right).$$

So, by Lemma 8.1,

$$\frac{\mu_{b_N}(A_{N,B-i}\boldsymbol{x}_{B-i}^* + A_{N,i}x_i')}{\mu_{b_N}(A_{N,B-i}\boldsymbol{x}_{B-i}^* + A_{N,i}x_i)} \geq e^{-\frac{\alpha\|A\|\left(\|A\|\left(\left\|x_{B-i}^*\right\|+1\right)+2\right)}{\sigma^2}} \geq e^{-\frac{1}{4(m+1)}} \geq 1-\frac{1}{4(m+1)},$$

by our choice of α.

We can also apply Lemma 8.3 to show that for each $j \notin N$,

$$
\frac{\Pr_{b_j}\left[b_j > A_{j,B-i}\boldsymbol{x}^*_{B-i} + A_{j,i}x'_i\right]}{\Pr_{b_j}\left[b_j > A_{j,B-i}\boldsymbol{x}^*_{B-i} + A_{j,i}x_i\right]}
$$

$$
\geq 1 - \frac{\alpha \left\|A\right\|\left(\left\|A\right\|\left(\left\|\boldsymbol{x}^*_{B-i}\right\|+1\right)\right)}{\sigma^2} e^{\frac{\alpha\|A\|\left(\|A\|\left(\left\|x^*_{B-i}\right\|+1\right)+3\right)}{\sigma^2}}
$$

$$
\geq 1 - \frac{2\alpha \left\|A\right\|\left(\left\|A\right\|\left(\left\|\boldsymbol{x}^*_{B-i}\right\|+1\right)\right)}{\sigma^2}, \qquad \text{as } e^{\frac{\alpha\|A\|\left(\|A\|\left(\left\|x^*_{B-i}\right\|+1\right)+3\right)}{\sigma^2}} \leq 2
$$

$$
\geq 1 - \frac{1}{2(m+1)},
$$

by our choice of α. Thus, we may conclude

$$
\frac{\rho(x'_i)}{\rho(x_i)} \geq \left(1 - \frac{1}{4(m+1)}\right)\left(1 - \frac{1}{2(m+1)}\right)^m \geq 1 - \frac{m+1}{2(m+1)} = 1/2.
$$

9. Connection to Smoothed Analysis of Simplex Method

The analysis of the simplex method in [ST01] is broken into two parts: a combinatorial bound on the smoothed size of a two-dimensional shadow of a linear program, and an analysis of a two-phase algorithm that uses this combinatorial bound as a black-box. The analysis of termination in this paper is closely related to the smoothed analysis of the shadow size. The intuition behind this analysis is that if the angle at a corner of the polytope of feasible points is bounded away from being flat, then the simplex method should make significant progress as it traverses this corner. The measure of angle used in [ST01] is approximately $\gamma(A, \boldsymbol{b}, \boldsymbol{c})$, at least for the corner optimizing the linear program. The size of the shadow, which upper bounds the number of steps taken by the simplex method, is then bounded by varying $\boldsymbol{c}$ over the plane onto which the shadow is projected.

The main technical lemma of the shadow-size analysis in [ST01], Lemma 4.0.11 (Angle bound), essentially says that for every $\boldsymbol{b}$, $\boldsymbol{c}$ and $\hat{A}$, the probability that a Gaussian perturbation A of $\hat{A}$ has $\alpha_D(A, \boldsymbol{b}, \boldsymbol{c})\gamma(A, \boldsymbol{b}, \boldsymbol{c}) < \epsilon$ is linear in ϵ, with a coefficient polynomial in n, m and σ. The most significant difference between this statement and the analysis in Lemma 6.2 and 6.4 is that in [ST01], $\boldsymbol{b}$ and $\boldsymbol{c}$ are not perturbed. This restriction seems necessary to apply the combinatorial bound in a black-box fashion in the analysis of the two-phase simplex algorithm. Also note that the simplex method analysis is for linear programs without the constraint $\boldsymbol{x} \geq 0$.

Otherwise, the arguments in this paper have a flavor very similar to those of [ST01], which mainly use the four techniques outlined in Section 8 of this paper; although, that paper uses more elaborate changes of variables. One probabilistic technique used in [ST01] that is absent in this paper is the Combination Lemma [ST01, Lemma 2.3.5] which allows one to obtain tight bounds on the

probability that a product of parameters is small from bounds on the probabilities that the individual parameters are small. The conditions of this lemma dictate the structure of the proofs in [ST01] as without it one could not obtain a bound on the probability of angle less than ϵ that is linear in ϵ. Moreover, without a bound that is linear in ϵ, one could not prove that the shadow has expected polynomial size. In contrast, in Lemma 6.5 of this paper the dependency is on $\epsilon^{1/3}$. It is possible that one could reduce this dependency using the combination lemma, but it is not essential for the results in this paper.

It is our hope that this paper will serve as a gentile introduction to the techniques used in the smoothed analysis of the simplex method.

10. Open Questions

By combining the smoothed analysis of Renegar's condition number from [DST02] and the smoothed analysis of termination from this work, we show that Renegar's and Ye's linear programming algorithms have smoothed complexity $O(m^3 \log(m/\sigma))$. That is, assuming the input linear program is subject to a Gaussian perturbation of variance σ^2, the expected running times of these algorithms is $O(m^3 \log(m/\sigma))$. We remark that this expectation averages over the cases in which the program is infeasible, unbounded, and bounded. It is not clear how each of these cases contributes to the expectation. It would be interesting to try to prove a similar result conditioning upon the program begin bounded and feasible. In this case, it would be natural to assume that the base program specified by $(\hat{A}, \hat{b}, \hat{c})$ is bounded and feasible as well. Similarly, if one wanted to analyze the running time conditioned upon the program being infeasible or unbounded, one should assume that the base program is as well. Such a result would not require modifying any of the proofs in this paper. However, the techniques of [DST02] would have to be improved to apply under such conditioning. As our analysis of the termination phase makes use of the result of [DST02] (Theorem 3.1 of this paper), this issue effects both the initialization and termination analyses.

We wish to conclude by observing that the termination phase described in this paper, and many of our definitions, only make sense for programs with unique optimal solutions. Under the model of perturbations used in this paper, the program has unique solutions with probability one, so this is not an issue. We ask whether one can define a model of perturbations of linear programs that results in degenerate linear programs, perhaps analogous to Todd's Degenerate Model [Tod91], and extend our results to this model.

References

[AC78]   David Avis and Vasek Chvátal. Notes on Bland's pivoting rule. In *Polyhedral Combinatorics*, volume 8 of *Math. Programming Study*, pages 24–34. 1978.

[AJPY93] K. M. Anstreicher, J. Ji, F. A. Potra, and Y. Ye. Average performance of a self–dual interior–point algorithm for linear programming. In P. M. Pardalos, editor, *Complexity in Numerical Optimization*, pages 1–15. World Scientific Publishing Co., London, United Kingdom, 1993.

[AJPY99] K. M. Anstreicher, J. Ji, F. A. Potra, and Y. Ye. Probabilistic analysis of an infeasible-interior-point algorithm for linear programming. *Mathematics of Operations Research*, 24(1):176–192, 1999.

[AKS87] I. Adler, R. M. Karp, and R. Shamir. A simplex variant solving an m x d linear program in $O(min(m^2, d^2))$ expected number of pivot steps. *J. Complexity*, 3:372–387, 1987.

[AM85]   Ilan Adler and Nimrod Megiddo. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM*, 32(4):871–895, October 1985.

[AZ99]   Nina Amenta and Gunter Ziegler. Deformed products and maximal shadows of polytopes. In B. Chazelle, J.E. Goodman, and R. Pollack, editors, *Advances in Discrete and Computational Geometry*, number 223 in Contemporary Mathematics, pages 57–90. Amer. Math. Soc., 1999.

[BD02]   Avrim Blum and John Dunagan. Smoothed analysis of the perceptron algorithm for linear programming. In *SODA '02*, pages 905–914, 2002.

[Bor77] Karl Heinz Borgwardt. *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung*. PhD thesis, Universitat Kaiserslautern, 1977.

[Bor80] Karl Heinz Borgwardt. *The Simplex Method: a probabilistic analysis*. Number 1 in Algorithms and Combinatorics. Springer-Verlag, 1980.

[BS95]   Avrim Blum and Joel Spencer. Coloring random and semi-random k-colorable graphs. *J. Algorithms*, 19(2):204–234, 1995.

[BV02]   Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 109–115. ACM Press, 2002.

[DST02] John Dunagan, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of interior point methods: Condition numbers. Available at `http://arxiv.org/abs/cs.DS/0302011`, 2002.

[EA96]   C. Mezaros X. Xu E. Andersen, J. Gondzio. Implementation of interior point methods for large scale linear programming. In T. Terlaky, editor, *Interior point methods in mathematical programming*. Kluwer Academic Publisher, 1996.

[Gol83] Donald Goldfarb. Worst case complexity of the shadow vertex simplex algorithm. Technical report, Columbia University, 1983.

[Gon88] Clovis C. Gonzaga. *An Algorithm for Solving Linear Programming Problems in $O(n^3 L)$ Operations*, pages 1–28. Springer-Verlag, 1988. Progress in Mathematical Programming, N. Megiddo ed..

[GS79]   Donald Goldfarb and William T. Sit. Worst case behaviour of the steepest edge simplex method. *Discrete Applied Math*, 1:277–285, 1979.

[Hai83] M. Haimovich. The simplex algorithm is very good!: On the expected number of pivot steps and related properties of random linear programs. Technical report, Columbia University, April 1983.

[IL94]   D. Shanno I. Lustig, R. Marsten. Interior point methods: computational state of the art. *ORSA Journal on Computing*, 6(1):1–14, 1994.

[Jer73] Robert G. Jeroslow. The simplex algorithm with the pivot rule of maximizing improvement criterion. *Discrete Math.*, 4:367–377, 1973.

[Kar84] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[Kha79] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademia Nauk SSSR*, pages 1093–1096, 1979.

[KM72]   V. Klee and G. J. Minty. How good is the simplex algorithm? In Shisha, O., editor, *Inequalities – III*, pages 159–175. Academic Press, 1972.

[LMS90]   I. J. Lustig, R. E. Marsten, and D. F. Shanno. The primal–dual interior point method on the Cray supercomputer. In T. F. Coleman and Y. Li, editors, *Large–Scale Numerical Optimization, Papers from the Workshop held at Cornell University, Ithaca, NY, USA, October 1989*, volume 46 of *SIAM Proceedings in Applied Mathematics*, pages 70–80. Society of Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 1990.

[Mur80]   K. G. Murty. Computational complexity of parametric linear programming. *Math. Programming*, 19:213–219, 1980.

[Ren88]   J. Renegar. A polynomial-time algorithm, based on Newton's method, for linear programming. *Math. Programming*, 40(1 (Ser. A)):59–93, 1988.

[Ren94]   J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994.

[Ren95a] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995.

[Ren95b] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Programming*, 70(3, Ser. A):279–351, 1995.

[Seg00]   Yoav Seginer. The expected norm of random matrices. *Combinatorics, Probability and Computing*, 9:149–166, 2000.

[Sma83]   S. Smale. On the average number of steps in the simplex method of linear programming. *Mathematical Programming*, 27:241–262, 1983.

[ST01]   Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. In *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing (STOC '01)*, pages 296–305, 2001. available at `http://arxiv.org/abs/cs.DS/0111050`.

[Tod86]   M.J. Todd. Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems. *Mathematical Programming*, 35:173–192, 1986.

[Tod91]   M. J. Todd. Probabilistic models for linear programming. *Mathematics of Operations Research*, 16(4):671–693, 1991.

[Tod94]   M. J. Todd. A lower bound on the number of iterations of primal-dual interior-point methods for linear programming. In G. A. Watson and D. F. Griffiths, editors, *Numerical Analysis 1993*, pages 237–259. Longman Press, Harlow, 1994.

[TY96]   M. J. Todd and Y. Ye. A lower bound on the number of iterations of long-step and polynomial interior-point methods for linear programming. *Annals of Operations Research*, 62:233–252, 1996.

[Vai90]   P. M. Vaidya. An algorithm for linear programming which requires $O((m+n)n^2 + (m+n)^{1.5}nL)$ arithmetic operations. *Mathematical Programming*, 47:175–201, 1990. Condensed version in: *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 29–38, 1987*.

[Wil65]   J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, 1965.

[Wri96]   S. Wright. *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, 1996.

[Ye92]   Yinyu Ye. On the finite convergence of interior-point algorithms for linear programming. *Mathematical Programming*, 57:325–335, 1992.

[Ye94]   Yinyu Ye. Toward probabilistic analysis of interior-point algorithms for linear programming. *Mathematics of Operations Research*, 19(1):38–52, 1994.

[Ye97]   Yinyu Ye. *Interior Point Algorithms*. Wiley-Interscience, 1997.