SAMPLING WITH REMOVAL IN LP-TYPE PROBLEMS

Bernd Gärtner

ABSTRACT. Random sampling is an important tool in optimization subject to finitely or infinitely many constraints. Here we are interested in obtaining solutions of low cost that violate only few constraints. Under convexity or similar favorable conditions, and assuming fixed dimension, one can indeed derive combinatorial bounds on the expected number (or probability mass) of constraints violated by the optimal solution subject to a (small) random sample of constraints. The cost of the sample solution, however, cannot be bounded combinatorially.

Suppose that after picking the sample, we remove \( k \) constraints from it in such a way that the best improvement of objective function is achieved. Can we still guarantee (good) bounds on the number or mass of violated constraints? This question was asked and partially answered in the framework of chance-constrained optimization [1].

Here we study the question in the completely combinatorial setting of LP-type problems, which allows a unified treatment of many different optimization problems. Given a nondegenerate LP-type problem of combinatorial dimension \( \delta \), we consider sampling with subsequent removal of the \( k \) constraints that lead to the best improvement of objective function. We show that this induces an LP-type problem of combinatorial dimension \( O(\delta^{k+1}) \), and that this bound is tight in the worst case. It follows that through this kind of removal, the expected number of violated constraints blows up by a factor of at most \( O(\delta^k) \); for relevant sample sizes, however, a matching lower bound for the blowup factor is not implied. At the same time, our results show that further improvements require new tools that go beyond combinatorial dimension.

1 Introduction

The combinatorial framework of LP-type problems is a well-studied abstraction of convex programming and encompasses many concrete (geometric) optimization problems. The framework has been introduced in a seminal paper by Matoušek, Sharir and Welzl in 1992 that also lists several concrete LP-type problems [12, 10]. Their randomized algorithm for solving LP-type problems yields—together with algorithms by Clarkson [4]—the currently best algorithm for linear programming in the real RAM model of computation [7].
Random sampling, the main tool in Clarkson’s algorithms, is well understood in this combinatorial framework: if we have an LP-type problem of (combinatorial) dimension $\delta$ over $n$ constraints, and we sample $r$ of them at random, then the expected number of constraints still violated by the sample solution is at most $\delta(n-r)/(r+1)$ [8].

In this paper, we obtain the first results for random sampling with removal in LP-type problems: if we sample $r$ constraints at random and then remove from the sample the $k$ ones that lead to the highest improvement in objective function value, the expected number of constraints violated by this “$k$-optimized” sample solution is $O(\delta^{k+1}(n-r)/(r+1))$ (assuming that $k$ is fixed). We also exhibit a concrete LP-type problem for which this bound is tight if $r = n-1$. In relevant cases, however (most notably in the aforementioned algorithms), we have $r \ll n$, and here we do not get a matching lower bound. We will discuss this issue (and the open problem arising from it) in detail in the last section.

The main technical step in the upper bound proof is the construction of a $k$-optimized LP-type problem of combinatorial dimension $O(\delta^{k+1})$, with the property that random sampling with removal in the original problem corresponds to “normal” random sampling in the $k$-optimized problem.

In practice, removing the “best $k$” constraints (the ones that lead to the best improvement in objective function value) is feasible only for very small $k$, and one would rather remove constraints according to some cheaper rule. It would be interesting to understand whether our combinatorial results extend to removal rules other than “the best $k$”. Currently, we do not know the answer. The “best $k$” assumption is most crucially used in Lemma 3 where we show that all removed constraints are violated by the improved solution. But also our construction of the $k$-optimized LP-type problem requires the “best $k$” rule (Lemma 2).

Still, our tight bound for “best-$k$-removal” has a clear theoretical value. On the one hand, for fixed $\delta$ and $k$, the expected number of violated constrains increases only by a constant factor, even under the best possible local improvement of the objective value. On the other hand, our lower bound shows that there is a price to pay that may be exponential in the number of removed constraints, unless new problem parameters that are finer than the combinatorial dimension can be found and employed.

### 1.1 Chance-constrained Optimization

The motivation for this work comes from chance-constrained optimization, where we have a probability distribution over a set of constraints. For a given $\varepsilon$, the goal is to find the best solution with violation probability at most $\varepsilon$, meaning that it violates a randomly chosen constraint with at most this probability. Let us call such a solution sufficiently feasible.

A natural approach here is to sample a finite number of constraints at random from the distribution, and to solve the problem optimally subject to the sampled constraints. Assuming a linear objective function and convex constraints, Campi and Garatti give tight bounds for the probability that the sample solution is sufficiently feasible, depending on the sample size [1]. In particular (and this was known before), the probability of not being sufficiently feasible tends to zero as the sample size goes to infinity.
In a later paper, Campi and Garatti study a variant of this approach where the sample solution is improved through removal of some of the sampled constraints, followed by reoptimization under the remaining ones [2]. In particular, if the quality of the sample solution suffers from outliers, this is desirable. The main question is how such a removal affects sufficient feasibility.

The main result of Campi and Garatti is that the probability of not being sufficiently feasible goes up, but only by a factor depending on the problem dimension $d$ and the number $k$ of removed constraints [2, Theorem 2.1]. As a consequence, the solution after removal is still sufficiently feasible with probability tending to one as the sample size goes to infinity.

It is not clear whether the bound of [2, Theorem 2.1] is best possible, and what happens if we are not interested in asymptotic results but results for a fixed budget of samples. Fagiano and Schildbach derive a bound for the expected violation probability of the sample solution but believe that this bound is poor [5].

Although the aforementioned work [1, 2, 5] deals with convex programming, many of the arguments are (implicitly) of a purely combinatorial nature and use little problem structure beyond what we also have in LP-type problems. This paper is an attempt to make the combinatorial properties of random sampling with removal explicit by working on the abstract level only. The author believes that the results obtained in this way will also contribute to the understanding of sampling with removal in chance-constrained optimization.

1.2 Paper Outline

In Section 2.2, we review the definition of LP-type problems and the known bounds for sampling without removal. Section 3 derives the $k$-optimized LP-type problem. In Section 4, we show how the known concept of $k$-levels in LP-type problems serves as a tool for the analysis of sampling with removal. Section 5 employs this tool to bound the combinatorial dimension of the $k$-optimized LP-type problem, from which we derive our main upper bound result in Section 6. Section 7 provides a matching lower bound on the combinatorial dimension. In the final Section 8, we discuss open questions, and in particular the limitations of our results.

2 Sampling in LP-type problems

On an abstract level, many problems of minimizing an objective function subject to finitely many constraints have the following structure.

Definition 1. Let $H$ be a finite set (the constraints), $\Omega$ a totally ordered set (the values) with a smallest element $-\infty$, and $w : 2^H \rightarrow \Omega$ a function that assigns an objective function value to each subset of constraints, such that $w(\emptyset) = -\infty$. Suppose that for all $F \subseteq G \subseteq H$, the following two axioms hold.

(i) $w(F) \leq w(G)$, and

(ii) If $w(F) = w(G) > -\infty$, then $w(G \cup \{h\}) > w(G) \Rightarrow w(F \cup \{h\}) > w(F)$. 


Then the triple $\mathcal{P} = (H, \Omega, w)$ is called an LP-type problem [12, 10]. We refer to axiom (i) as monotonicity and axiom (ii) as locality.

In an LP-type problem, $w(G)$ represents the minimum objective function value subject to only the constraints in $G$. The original minimization problem asks for the value $w(H)$, the minimum value subject to all constraints. The locality axiom can be seen as an abstract analog of convexity of constraints, since it allows for a local test of optimality: a simple induction shows that if $w(F \cup \{h\}) = w(F) > -\infty$ for all $h \in G \setminus F$, then $w(F) = w(G)$.

2.1 Smallest Enclosing Balls (SEB)

As an example, we use the classic problem of computing the smallest enclosing ball of a finite set of points $P$ in Euclidean space $\mathbb{R}^d$. To write this as an LP-type problem, we let $H = P$ and $\Omega = \mathbb{R} \cup \{-\infty\}$. For $G \subseteq H$, we define $w(G)$ to be the radius of the smallest enclosing ball of $G$, with the convention that the empty set yields radius $-\infty$. The smallest enclosing ball of a nonempty set of points exists and is unique [13], hence $w$ is well-defined. Monotonicity is clear, and to see locality, we observe that if $F \subseteq G$, $w(F) = w(G)$ means that both $F$ and $G$ have the same smallest enclosing ball. If $w(G \cup \{h\}) > w(G)$, then $h$ is outside of that ball, so we also have $w(F \cup \{h\}) > w(F)$. Let us refer to this LP-type problem as SEB.

2.2 The Sampling Lemma

Given a minimization problem subject to a finite (but possibly very large) number of constraints, one approach for obtaining an approximate solution is the following: sample a (much smaller) subset of constraints at random, and optimally solve the problem subject to only the sampled constraints. In the framework of LP-type problems, we can define a “combinatorial quality” of the sample solution, by counting the number of constraints still violated by it.

**Definition 2.** Let $\mathcal{P} = (H, \Omega, w)$ be an LP-type problem, $G \subseteq H$.

(i) A constraint $h \in H \setminus G$ is *violated* by $G$ if $w(G \cup \{h\}) > w(G)$. We use $V_{\mathcal{P}}(G)$ to denote the set of constraints violated by $G$.

(ii) A constraint $h \in G$ is *extreme* in $G$ if $w(G \setminus \{h\}) < w(G)$. We use $X_{\mathcal{P}}(G)$ to denote the set of extreme constraints in $G$.

Let us illustrate these notions for SEB. As already seen, $h$ violates $G$ if $h$ is outside of the smallest enclosing ball of $G$; $h$ is extreme in $G$ if its removal allows for a smaller enclosing ball. Necessarily, such an $h$ is on the boundary of the smallest enclosing ball, but this is not sufficient: if $G$ consists of the four corners of a square, then $G$ has no extreme points.
Lemma 1 (Sampling Lemma [8]). Let $\mathcal{P} = (H, \Omega, w)$ be an LP-type problem, and $r \in \mathbb{N}$, $0 \leq r < n := |H|$. Let $R$ be a random sample of $H$ of size $r$, obtained by choosing uniformly at random from all $r$-element subsets of $H$. Let $v_r := E[|V_P(R)|]$ be the expected number of constraints violated by $R$, and let $x_r := E[|X_P(R)|]$ denote the expected number of extreme constraints in $R$. Then

$$v_r = \frac{n - r}{r + 1} x_r + 1.$$ 

The Sampling lemma can be used to argue that $v_r$ is small if the expected number $x_r$ of extreme constraints in a random sample of size $r$ is small. For smallest enclosing balls in $\mathbb{R}^d$, this is the case: using Helly’s Theorem, one can show that every set has at most $d + 1$ extreme points, and this yields $v_r \leq (d + 1)(n - r)/(r + 1)$. For example, if $d = 2$, the smallest enclosing ball of a random sample of size $\sqrt{n}$ has in expectation less than $3\sqrt{n}$ points outside.

3 The $k$-optimized Problem

Suppose we are given an LP-type problem $\mathcal{P} = (H, \Omega, w)$ and a random sample of constraints $R \subseteq H$. We have seen that under favorable conditions, Lemma 1 provides a good bound on the expected number of constraints still violated by $R$. We now want to understand how this number changes when we remove some fixed number $k$ of constraints from the sample in such a way that the best improvement in objective function value is obtained.

Definition 3. Let $\mathcal{P} = (H, \Omega, w)$ be an LP-type problem, and let $k \in \mathbb{N}$. We define the function $w_k : 2^H \rightarrow \Omega$ by

$$w_k(G) := \min_{C \subseteq G, |G \setminus C| \leq k} w(C), \quad G \subseteq H.$$ 

A subset $C$ of $G$ such that $|G \setminus C| \leq k$ is a $k$-candidate of $G$. A $k$-candidate attaining the minimum in (1) is called a $k$-witness of $G$.

This means, we are interested in the smallest value that can be obtained by removing at most $k$ constraints from $G$. Note that $w_0 = w$. It turns out that $w_k$ inherits monotonicity from $w$.

Lemma 2. Let $\mathcal{P} = (H, \Omega, w)$ be an LP-type problem, and let $k \in \mathbb{N}$. For $F \subseteq G$, we have $w_k(F) \leq w_k(G)$.

Proof. Let $C$ be a $k$-witness of $G$, and define $C' = C \cap F$. We have $F \setminus C' = F \setminus C \subseteq G \setminus C$, hence $|F \setminus C'| \leq k$, and $C'$ is a $k$-candidate of $F$. This implies that

$$w_k(F) \leq w(C') \leq w(C) = w_k(G),$$

where we have used monotonicity of $w$ in the second inequality.

Definition 4. If $\mathcal{P} = (H, \Omega, w)$ is an LP-type problem and $k \in \mathbb{N}$, then $\mathcal{P}_k = (H, \Omega, w_k)$ is the $k$-optimized version of $\mathcal{P}$.

Without additional assumptions (to be introduced next), $\mathcal{P}_k$ is not an LP-type problem.
3.1 Weak Nondegeneracy

We next want to argue that every $k$-witness $C$ of $G$ violates all removed constraints $h \in G \setminus C$. But this is not true in general (consider an LP-type problem that assigns the same value to every nonempty subsets of constraints). We therefore need to make an additional assumption.

**Definition 5.** An LP-type problem $P = (H, \Omega, w)$ is called weakly nondegenerate if every set $G$ with $w(G) > -\infty$ has at least one extreme constraint.

For example, for SEB in the plane, the set of 4 corners of a square would be a set without any extreme points. If no four points are cocircular, however, we always obtain a weakly nondegenerate instance.

**Lemma 3.** Let $P = (H, \Omega, w)$ be a weakly nondegenerate LP-type problem, $k \in \mathbb{N}$. Let $G \subseteq H$ such that $w_k(G) > -\infty$, and let $C$ be a $k$-witness of $G$. Then the following two statements hold.

(i) $|G \setminus C| = k$.

(ii) $V_P(C) \supseteq G \setminus C$.

**Proof.** (i) Because of $w(C) = w_k(G) > -\infty$, $C$ has an extreme constraint $x \in C$ with $w(C \setminus \{x\}) < w(C)$. This implies $|G \setminus C| = k$, since otherwise, $C \setminus \{x\}$ would be a $k$-candidate of $G$ with value smaller than $w(C)$—contradiction.

(ii) Suppose that $w(C \cup \{h\}) = w(C)$ for some $h \in G \setminus C$. Since $w(C \cup \{h\}) = w(C) = w_k(G) > -\infty$, the set $C \cup \{h\}$ has an extreme constraint $h' \neq h$, meaning that $D := C \cup \{h\} \setminus \{h'\}$ satisfies $|D| = |C|$ and is another $k$-candidate of $G$. But $h'$ is extreme in $C \cup \{h\}$, so $w(D) < w(C \cup \{h\}) = w(C)$, a contradiction to $C$ being a $k$-witness of $G$. $\square$

The following is the main technical lemma of this section. Recall that we want to count the constraints violated by a random sample $R$ in $P$, after removal of the “best $k$” constraints from $R$. Assume for a moment that $k$-witnesses in $P$ are unique (we will take care of this afterwards). Then the lemma says that on top of the $k$ removed ones (previous lemma), the constraints in question are precisely the ones violated by $R$ in the $k$-optimized version $P_k$. Thus, we can reduce random sampling with removal in $P$ to the well-understood normal random sampling in $P_k$.

**Lemma 4.** Let $P = (H, \Omega, w)$ be a weakly nondegenerate LP-type problem, $k \in \mathbb{N}$, and $G \subseteq H$ with $w_k(G) > -\infty$. Then the following two statements are equivalent for every $h \in H \setminus G$.

(i) $w_k(G \cup \{h\}) = w_k(G)$.

(ii) There is a $k$-witness $C$ of $G$ with $w(C \cup \{h\}) = w(C)$. 
Proof. Let us first show (ii) ⇒ (i). Let \( C \) be a \( k \)-witness of \( G \). We have \(|(G \cup \{h\}) \setminus (C \cup \{h\})| = |G \setminus C| \leq k\), so \( C \cup \{h\} \) is a \( k \)-candidate of \( G \cup \{h\} \). If \( w(C \cup \{h\}) = w(C) \), we therefore get

\[
w_k(G \cup \{h\}) \leq w(C \cup \{h\}) = w(C) = w_k(G),
\]

and monotonicity of \( w_k \) (Lemma 2) yields \( w_k(G \cup \{h\}) = w_k(G) \).

For the direction (i) ⇒ (ii), we have \( w_k(G \cup \{h\}) = w_k(G) \), and we let \( D \) be a \( k \)-witness of \( G \cup \{h\} \). We will show that

(a) \( h \in D \),

(b) \( C := D \setminus \{h\} \) is a \( k \)-witness of \( G \) satisfying \( w(C \cup \{h\}) = w(C) \).

To see (a), suppose \( h \notin D \), so \( D \subseteq G \) and \( |G \setminus D| = k - 1 \). Then \( D \setminus \{x\} \) is a \( k \)-candidate of \( G \) for all \( x \in D \), meaning that

\[
w_k(G) \leq w(D \setminus \{x\}) \leq w(D) = w_k(G \cup \{h\}) = w_k(G),
\]

so we have equality throughout, and \( D \) is a set of value \( w_k(G) > -\infty \) without extreme constraints. This is a contradiction to weak nondegeneracy of \( \mathcal{P} \).

For (b), we use \( h \in D \) to conclude that \( |G \setminus (D \setminus \{h\})| = |(G \cup \{h\}) \setminus D| \leq k \), so \( C := D \setminus \{h\} \) is a \( k \)-candidate of \( G \). Then we argue as before that

\[
w_k(G) \leq w(C) \leq w(C \cup \{h\}) = w(D) = w_k(G \cup \{h\}) = w_k(G),
\]

so we have equality throughout and \( C \) is a \( k \)-witness of \( G \) as required. \( \square \)

Already in the SEB problem, we can see that the assumption of weak nondegeneracy is necessary for the previous lemma to hold. Let \( G \) be the four corners of a square, and let \( h \) be another point far away from the square. We have \( w_1(G \cup \{h\}) = w_1(G) \); in both cases, the smallest disk that can be obtained by removing one point is the smallest enclosing disk of the square. However, there is no \( 1 \)-witness \( C \) of \( G \) such that \( w(C \cup \{h\}) = w(C) \): indeed, \( h \) is violated by all subsets of \( G \).

### 3.2 Strong Nondegeneracy

Under weak nondegeneracy, a set may still have several \( k \)-witnesses, but we must avoid this in order to complete the argument outlined before Lemma 4. We therefore stipulate a stronger notion of nondegeneracy.

**Definition 6.** Let \( \mathcal{P} = (H, \Omega, w) \) be an LP-type problem.

(i) A **basis** is a set \( B \subseteq H \) with \( X_\mathcal{P}(B) = B \); in other words \( w(B \setminus \{h\}) < w(B) \) for all \( h \in B \).

(ii) A **basis of** \( G \subseteq H \) is a basis \( B \subseteq G \) such that \( w(B) = w(G) \).
(iii) \( \mathcal{P} \) is called strongly nondegenerate if for every two bases \( B, B' \subseteq H \), \( w(B) = w(B') \) implies \( B = B' \).

We remark that any inclusion-minimal subset of \( G \) with value \( w(G) \) is a basis of \( G \), and vice versa. Under strong nondegeneracy, every set \( G \) has exactly one basis.

**Lemma 5.** A strongly nondegenerate LP-type problem is weakly nondegenerate.

**Proof.** Suppose \( \mathcal{P} = (H, \Omega, w) \) is strongly nondegenerate and let \( G \subseteq H \) be a set such that \( w(G) > -\infty \). We need to show that \( G \) has an extreme constraint. Suppose not and let \( B \subseteq G \) be a basis of \( G \).

Because of \( w(G) = w(\emptyset) \), \( B \) contains at least one constraint \( h \). Since \( h \) is not extreme in \( G \), we have \( w(G) = w(G \setminus \{ h \}) \), and every basis \( B' \) of \( G \setminus \{ h \} \) is also a basis of \( G \). But since \( h \notin B' \), we have two distinct bases \( B, B' \) with the same value \( w(G) \), a contradiction to strong nondegeneracy.

As an illustration, consider SEB again. A basis of \( G \) is a minimal subset of points with the same smallest enclosing ball of \( G \). In particular, all points of the basis are on the ball’s boundary. The set of four corners of a square has two bases: the two pairs of diagonally opposite points.

Maybe surprisingly, SEB is not strongly nondegenerate, even if we assume sufficiently general position of the points. Indeed, every basis of size 1 has the same value 0 (the radius of the smallest enclosing ball of a single point). However, we can define new values \( w'(G) = (w(G), V_{\text{SEB}}(G)) \). We compare these pairs lexicographically, where we use a fixed but arbitrary order among sets in the second component. With respect to \( w' \), we have strong nondegeneracy if no \( d + 2 \) points are cospherical. We later come back to this refinement [9] in a more general setting.

The following corollary of the above definition shows that under strong nondegeneracy, sets with the same value violate the same constraints, and this will be the key to proving uniqueness of \( k \)-witnesses below.

**Corollary 1.** Let \( \mathcal{P} = (H, \Omega, w) \) be a strongly nondegenerate LP-type problem. For all \( C, C' \subseteq H \), \( w(C) = w(C') \) implies \( V_{\mathcal{P}}(C) = V_{\mathcal{P}}(C') \).

**Proof.** Let \( B \) and \( B' \) be bases of \( C \) and \( C' \), respectively. Since \( w(C) = w(C') \), we also have \( w(B) = w(B') \), so \( B = B' \) by strong nondegeneracy. Locality then yields \( V_{\mathcal{P}}(B) = V_{\mathcal{P}}(C) = V_{\mathcal{P}}(C') \). \( \square \)

Under strong nondegeneracy, we can now strengthen our previous “workhorse” Lemma 4.

**Lemma 6.** Let \( \mathcal{P} = (H, \Omega, w) \) be a strongly nondegenerate LP-type problem, \( k \in \mathbb{N} \), and \( G \subseteq H \) with \( w_k(G) > -\infty \). Then the following two statements hold.

(i) \( G \) has a unique \( k \)-witness, which we denote by \( C_k(G) \).

(ii) For \( h \in H \setminus G \),

\[
\begin{align*}
    w_k(G \cup \{ h \}) > w_k(G) & \iff w_k(C_k(G) \cup \{ h \}) > w_k(C_k(G)).
\end{align*}
\]
Proof. For (i), suppose that there are two \( k \)-witnesses \( C, C' \). By Corollary 1, \( V_P(C) = V_P(C') \), and in particular \( V_P(C) \cap G = V_P(C') \cap G \). On the other hand, Lemma 3 yields \( V_P(C) \cap G = G' \setminus C \) and \( V_P(C') \cap G = G' \setminus C' \), so \( C = C' \) follows. Given that \( G \) has a unique \( k \)-witness, part (ii) is an easy consequence of Lemma 4.

We conclude this section by showing that the \( k \)-optimized problem is in fact an LP-type problem if we assume strong nondegeneracy.

**Theorem 1.** Let \( P = (H, \Omega, w) \) be a strongly nondegenerate LP-type problem, \( k \in \mathbb{N} \). Then the \( k \)-optimized version \( P_k = (H, \Omega, w_k) \) (Definition 4) is an LP-type problem.

Proof. Monotonicity is Lemma 2. For locality, fix \( F \subseteq G \) such that \( w_k(F) = w_k(G) = w(C_k(F)) = w(C_k(G)) > -\infty \) and \( h \in H \setminus G \) such that \( w_k(G \cup \{h\}) > w_k(G) \). By Lemma 6(ii), \( w(C_k(F) \cup \{h\}) > w(C_k(G)) \), and from Corollary 1, we get \( V_P(C_k(F)) = V_P(C_k(G)) \). So we also have \( w(C_k(F) \cup \{h\}) > w(C_k(F)) \). Applying Lemma 6(ii) again, this time to \( F \), we get \( w_k(F \cup \{h\}) > w_k(F) \) and hence locality.

4 The \( k \)-level

We have now shown that the \( k \)-optimized version \( P_k \) of a strongly nondegenerate LP-type problem \( P \) is an LP-type problem itself, and the next goal is to apply the Sampling Lemma 1 to get bounds on the expected number of constraints violated by the solution of a random sample in \( P_k \). But for this, we need to bound the (expected) number of extreme constraints in a random sample of size \( r + 1 \). This is done using \( k \)-levels.

**Definition 7.** Let \( P = (H, \Omega, w) \) be a strongly nondegenerate LP-type problem, \( G \subseteq H, k \in \mathbb{N} \) such that \( w_k(G) > -\infty \). The \( k \)-level of \( G \) is the set

\[
\mathcal{L}_k^P(G) = \{ C \subseteq G : |G \setminus C| = k, \ V_P(C) \supseteq G \setminus C \}.
\]

According to Lemma 3, the \( k \)-level of \( G \) contains the \( k \)-witness \( C_k(G) \) of \( G \) and possibly many more sets. The crucial fact we prove next is that the size of the \((k+1)\)-level of \( G \) provides a bound for the number of extreme constraints of \( G \) w.r.t. the \( k \)-optimized problem \( P_k \).

**Theorem 2.** Let \( P = (H, \Omega, w) \) be a strongly nondegenerate LP-type problem, \( G \subseteq H, k \in \mathbb{N} \) such that \( w_{k+1}(G) > -\infty \). Then we have

\[
|X_{P_k}(G)| \leq (k+1)|\mathcal{L}_{k+1}^P(G)|.
\]

Proof. Let \( x \in X_{P_k}(G) \) (meaning that \( w_k(G \setminus \{x\}) < w_k(G) \)), and let \( C_x \) be the unique \( k \)-witness of \( G \setminus \{x\} \). Since \( w_{k+1}(G) > -\infty \) implies \( w_k(G \setminus \{x\}) > -\infty \), Lemma 3 yields

(i) \( |G \setminus C_x| = k + 1 \),

(ii) \( V_P(C_x) \supseteq (G \setminus \{x\}) \setminus C_x \).
But we also have \( x \in V_P(C_x) \) as a consequence of Lemma 6 (ii). This implies that \( C_x \in L_{k+1}^P(G) \), and we will charge \( x \) to \( C_x \). Taken over all \( x \in X_{P_k}(G) \), every set \( C \) in the \((k+1)\)-level is charged at most \( k+1 \) times (namely at most once for every \( x \in G \setminus C \)). The claimed bound follows.

\[ \]

5 Combinatorial Dimension

In a general LP-type problem, we cannot make any nontrivial statements about the size of a level according to Definition 7. But in many applications, the LP-type problem has fixed dimension.

**Definition 8.** Let \( P = (H, \Omega, w) \) be an LP-type problem. The **combinatorial dimension** \( \delta(P) \) of \( P \) is the size of a largest basis.

There is a strong relation between combinatorial dimension and extreme constraints.

**Lemma 7.** Let \( P = (H, \Omega, w) \) be an LP-type problem of combinatorial dimension \( \delta \). Then \( |X_P(G)| \leq \delta \) for all \( G \subseteq H \), and there exists a basis \( B \subseteq H \) such that \( |X_P(B)| = \delta \).

**Proof.** The latter statement immediately follows from Definition 6 of a (largest) basis. For the former, we use the simple observation that an extreme constraint of \( G \) is contained in every basis of \( G \).

For the smallest enclosing ball problem SEB in \( \mathbb{R}^d \), the combinatorial dimension is thus at most \( d+1 \), a consequence of the fact that there are always at most \( d+1 \) extreme points (Section 2.2).

For strongly nondegenerate LP-type problems with fixed combinatorial dimension, we have complete control over the level sizes, and this is a classic result first proved by Clarkson in the context of linear programming [3], and later by Matoušek for LP-type problems [9]; below we apply yet another version [8, Theorem 4.1]. In fact, these results only require “normal” nondegeneracy.

**Definition 9.** Let \( P = (H, \Omega, w) \) be an LP-type problem. \( P \) is called **nondegenerate** if every set \( G \subseteq H \) has a unique basis.

**Lemma 8.** Every strongly nondegenerate LP-type problem is nondegenerate, and every nondegenerate LP-type problem is weakly nondegenerate.

**Proof.** Nondegeneracy from strong nondegeneracy: if a set has two bases \( B, B' \), these have the same value, hence \( B = B' \) by strong nondegeneracy. Weak nondegeneracy from nondegeneracy: Let \( G \) be a set with \( w(G) > -\infty \). Let \( B \subseteq G \) be the unique basis of \( G \). Because of \( w(G) = w(B) > -\infty \), we have \( B \neq \emptyset \), and every constraint \( x \in B \) is extreme in \( G \), since otherwise, we would find another basis in \( G \setminus \{x\} \).

It is important to mention that every nondegenerate LP-type problem can be refined to a strongly nondegenerate one of the same combinatorial dimension. We have explained
how this works for SEB after Definition 6, and the general construction is due to Matoušek [9]. A refinement $P'$ of $P$ has the property that

$$V_P(G) \subseteq V_{P'}(G)$$

for all $G \subseteq H$, so the refinement can only add violated constraints. As a consequence of this discussion, a nondegenerate problem can be assumed to be strongly nondegenerate without loss of generality. Here is the bound on the $k$-level.

**Theorem 3.** Let $P = (H, \Omega, w)$ be a nondegenerate LP-type problem of combinatorial dimension $\delta$, and let $G \subseteq H, k \in \mathbb{N}$ such that $w_k(G) > -\infty$ and $|G| \geq k + \delta$. Then

$$|L^P_k(G)| \leq \binom{k + \delta - 1}{\delta - 1}.$$ 

**Proof.** We apply a further refinement [9, Lemma 2.4] of $P$ to a regular nondegenerate LP-type problem $P'$, meaning that every set $G$ of size at least $\delta$ has a unique basis of size exactly $\delta$. Then we have

$$|L^P_k(G)| = \binom{k + \delta - 1}{\delta - 1},$$

by the aforementioned classic results [8, Theorem 4.1]. Property (2) yields $L^P_k(G) \subseteq L^P_{k+1}(G)$, since there can never be more than $k$ violated constraints. The statement of the theorem follows. \hfill $\Box$

## 6 Sampling with Removal

We are approaching the main result of this paper that allows us to bound the expected number of constraints violated by a random sample $R$ in an LP-type problem, after removal of the “best $k$” constraints from $R$. We start with the combinatorial dimension of the $k$-optimized version.

**Lemma 9.** Let $P = (H, \Omega, w)$ be a strongly nondegenerate LP-type problem of combinatorial dimension $\delta$, let $k \in \mathbb{N}$, and let $P_k = (H, \Omega, w_k)$ be its $k$-optimized version (which is an LP-type problem by Theorem 1). Then $P_k$ has combinatorial dimension at most

$$(k + 1) \binom{k + \delta}{\delta - 1} = O \left( \frac{k + \delta}{\delta - 1} \right).$$

for fixed $k$.

**Proof.** We already know that every set $G$ has at most

$$|X_{P_k}(G)| \leq (k + 1) |L^{P_k}_k(G)| \leq (k + 1) \binom{k + \delta}{\delta - 1}$$

extreme constraints, by the bounds established in Theorems 2 and 3. According to Lemma 7, this also bounds the combinatorial dimension. \hfill $\Box$

Here is our main result.
Theorem 4. Let $\mathcal{P} = (H, \Omega, w)$ be a strongly nondegenerate LP-type problem of combinatorial dimension $\delta$, and let $r, k \in \mathbb{N}, 0 \leq k + \delta \leq r < n$. Let $R \subseteq H$ be a set of $r$ constraints chosen uniformly at random, and let $C_k(R)$ be the unique $k$-witness of $R$, i.e. an $(r - k)$-subset of $R$ with minimum possible $w$-value.

Let $v_{k,r} = E[|V_{\mathcal{P}}(C_k(R))|]$ be the expected number of constraints violated by $C_k(R)$. Then

$$v_{k,r} \leq k + \frac{n - r}{r + 1} \binom{k + \delta}{\delta - 1} = O\left(\delta k + 1 \frac{n - r}{r + 1}\right),$$

for fixed $k$.

Proof. Lemmas 3 and 6 together show that there are two types of constraints violated by $C_k(R)$ w.r.t. the original LP-type problem $\mathcal{P}$: On the one hand, the $k$ constraints in $R \setminus C_k(R)$ (Lemma 3), and on the other hand the constraints in $H \setminus R$ that are violated by $R$ w.r.t. the $k$-optimized LP-type problem $\mathcal{P}_k$ (Lemma 6). Taking expectations, this yields $v_{k,r} = k + E[|V_{\mathcal{P}}(R)|]$, and the Sampling Lemma 1 (applied to the $k$-optimized version $\mathcal{P}_k$) yields

$$E[|V_{\mathcal{P}_k}(R)|] \leq \frac{n - r}{r + 1} \delta(\mathcal{P}_k),$$

since $\delta(\mathcal{P}_k)$ is by Lemma 7 an upper bound for the (expected) number of extreme constraints of any set. Plugging in the bound for $\delta(\mathcal{P}_k)$ from Lemma 9, the result follows. \[\square\]

7 Lower Bound

The quality of the bound in Theorem 4 is determined by the combinatorial dimension of the $k$-optimized LP-type problem, and it is natural to ask whether the bound that we have given in Lemma 9 is asymptotically best possible for the relevant case of fixed $k$. We currently have no “geometric” LP-type problem (such as SEB) for which we can prove tightness of the bound, but there is an abstract LP-type problem that does it.

7.1 The Top Nodes LP-type Problem

Let $H$ be the set of nodes of a rooted tree, and let $C(u)$ be the set of children of node $u$. We assign positive weights to the nodes such that

$$w(u) > \sum_{v \in C(u)} w(v)$$

for every node $u$. For $G \subseteq H$, we let $w(G) = \sum_{u \in G} w(u)$ be the total weight of $G$. We choose $w$ generic, meaning that different sets have different total weights.

For a set of nodes $G$, its top nodes are the nodes that have no ancestor in $G$, and we use $T(G) \subseteq G$ to denote the set of top nodes in $G$. For example, any set $G$ containing the root node $\rho$ satisfies $T(G) = \{\rho\}$. 
Definition 10. For a fixed positive integer $\delta$, let $w^{(\delta)} : 2^H \to \mathbb{R}_{\geq 0}$ be defined as follows for every set $G \subseteq H$.

$$w^{(\delta)}(G) := \max_{B \subseteq T(G),|B| \leq \delta} w(B).$$

Let $B(G)$ be the unique set of at most $\delta$ top nodes that achieves the maximum in (4); uniqueness follows from $w$ being generic.

We have $B(\emptyset) = \emptyset$ and hence $w^{(\delta)}(\emptyset) = 0$. All other sets have at least one top node and therefore positive value.

Theorem 5. Problem $\mathcal{T} = (H, R, w^{(\delta)})$ is a strongly nondegenerate LP-type problem of combinatorial dimension at most $\delta$.

To prove this, we separately show monotonicity and locality (Lemmas 11 and 13 below). The bound on the combinatorial dimension (Definition 8) is a consequence of (4), in form of the simple observation that $w^{(\delta)}(G) = w^{(\delta)}(B(G))$. Strong nondegeneracy (Definition 6) immediately follows from $w$ being generic. The reader mainly interested in the lower bound construction can skip ahead to Section 7.2.

Here is a preparatory lemma.

Lemma 10. Let $G \subseteq H$, and let $S$ be a set of top nodes in $G$ with a common ancestor $h \notin G$. Then $w(h) > w(S)$.

Proof. We can obtain $S$ from the set $\{h\}$ by repeatedly replacing a node with some of its children. In each such replacement, the weight of the current set decreases according to (3). \hfill \square

Now we can show monotonicity of $w^{(\delta)}$.

Lemma 11. For all $G \subseteq H, h \in H \setminus G$, we have

$$w^{(\delta)}(G) \leq w^{(\delta)}(G \cup \{h\}).$$

Proof. If $h$ is not a top node in $G \cup \{h\}$, we have $T(G \cup \{h\}) = T(G)$ and hence $w^{(\delta)}(G) = w^{(\delta)}(G \cup \{h\})$. Otherwise, let $Q \subseteq G$ be the (possibly empty) set of nodes of $G$ which have $h$ as an ancestor. We thus have

$$T(G \cup \{h\}) = (T(G) \setminus Q) \cup \{h\}.$$

If $B(G) \cap Q = \emptyset$, we have $B(G) \subseteq T(G \cup \{h\})$, meaning that

$$w^{(\delta)}(G \cup \{h\}) \geq w(B(G)) = w^{(\delta)}(G),$$

so the statement of the lemma holds. If $B(G) \cap Q \neq \emptyset$, we define

$$B' = (B(G) \setminus Q) \cup \{h\} \subseteq T(G \cup \{h\}).$$
We get $|B'| \leq |B(G)| \leq \delta$ and

$$w(B') = w(B(G)) + w(h) - w(B(G) \cap Q) > w(B(G)),$$

where the latter inequality uses Lemma 10 applied with the set $S = B(G) \cap Q$ of top nodes. Hence,

$$w^{(\delta)}(G \cup \{h\}) \geq w(B') > w(B(G)) = w^{(\delta)}(G),$$

and monotonicity is established. \qed

Here is the major step towards locality: Whether a constraint $h$ is violated by $G$ only depends on $B(G)$.

**Lemma 12.** Let $G \subseteq H$, and let $h \in H \setminus G$ be a top node in $G \cup \{h\}$. Then the following two statements are equivalent.

(i) $w^{(\delta)}(G) < w^{(\delta)}(G \cup \{h\})$.

(ii) $|B(G)| < \delta$, or $w(h) > \min_{u \in B(G)} w(u)$.

**Proof.** We start with the direction (ii)$\Rightarrow$(i). Let $Q \subseteq B(G)$ be the set of nodes in $B(G)$ with ancestor $h$. If $Q \neq \emptyset$, then $B' = (B(G) \setminus Q) \cup \{h\}$ is a set of at most $\delta$ top nodes in $G \cup \{h\}$, proving that $w^{(\delta)}(G \cup \{h\}) \geq w(B') > w(B(G)) = w^{(\delta)}(G)$, where the strict inequality follows from Lemma 10. If $Q = \emptyset$, the previous argument still works with the same $B'$ in the case $|B(G)| < \delta$, and if $w(h) > \min_{u \in B(G)} w(u)$, it works with $B' = B(G) \cup \{h\} \setminus \{\arg\min_{u \in B(G)} w(u)\}$.

For the direction $\neg$(ii)$\Rightarrow\neg$(i), let $B' = B(G \cup \{h\})$. If $h \notin B'$, we have $B' \subseteq T(G \cup \{h\}) \setminus \{h\} \subseteq T(G)$, hence $w^{(\delta)}(G) \geq w(B') = w^{(\delta)}(G \cup \{h\})$. So we can assume that $h \in B'$ and argue as follows: Since $|B(G)| = \delta$, we have a node $\tilde{u} \in B(G) \setminus B'$, and since $w(h) \leq \min_{u \in B(G)} w(u) \leq w(\tilde{u})$, the set $B'' = B' \cup \{\tilde{u}\} \setminus \{h\}$ is a set of at most $\delta$ top nodes in $G$ satisfying $w^{(\delta)}(G) \geq w(B'') \geq w(B') = w^{(\delta)}(G \cup \{h\})$. \qed

Locality is now an easy consequence.

**Lemma 13.** Let $F \subseteq G \subseteq H$ with $w^{(\delta)}(F) = w^{(\delta)}(G)$, and let $h \in H \setminus G$. If $w^{(\delta)}(G) < w^{(\delta)}(G \cup \{h\})$, then also $w^{(\delta)}(F) < w^{(\delta)}(F \cup \{h\})$.

**Proof.** If $w^{(\delta)}(G) < w^{(\delta)}(G \cup \{h\})$, then $h$ is necessarily a top node in $G \cup \{h\}$ and therefore also in $F \cup \{h\}$. Moreover, by genericity of $w$, $w^{(\delta)}(F) = w^{(\delta)}(G)$ implies $B(F) = B(G)$. Then the statement follows from Lemma 12. \qed

### 7.2 Lower Bound Construction

We now show that for a suitable tree and suitable node weights, the $k$-optimized LP-type problem $T_k = (H, R, w_k^{(\delta)})$ derived from $T$ has combinatorial dimension $\Omega(\delta^{k+1})$. 
Let us assume that \( k \geq 1 \) is a constant, and that \( k + 1 \) divides \( \delta \geq 2 \). Then our tree will be the complete \( \delta/(k+1) \)-ary tree with \( k+3 \) levels. The total number of nodes is therefore
\[
n := |H| = \sum_{t=0}^{k+2} \left( \frac{\delta}{k+1} \right)^t = \Theta \left( \delta^{k+2} \right).
\]

Here is the idea of the lower bound construction: To show that \( T_k \) has combinatorial dimension \( \Omega(\delta^{k+1}) \), we exhibit a set \( G \subseteq H \) with that many extreme constraints. We choose \( G = H \setminus \{\rho\} \), i.e. the set of all nodes except the root, and we show that all interior nodes in \( G \) are extreme. Since the number of such interior nodes is
\[
\sum_{t=1}^{k+1} \left( \frac{\delta}{k+1} \right)^t = \Theta \left( \delta^{k+1} \right),
\]
this yields the desired bound. To make all the interior nodes extreme, we choose the weight function \( w \) in such a way that \( w_k^\delta(G) > w_k^\delta(G \setminus \{h\}) \) for every interior node \( h \). What makes this possible is the height of the tree: starting from \( G \), the removal of \( k \) nodes cannot turn any leaf into a top node, but starting from \( G \setminus \{h\} \), this is always possible. To exploit this, we need a weight function that assigns exceptionally small values to the leaves; then the most profitable way to remove \( k \) constraints from \( G \setminus \{h\} \) is to expose some leaf as a top node.

We choose \( w(u) \in (0, 1) \) if \( u \) is a leaf, and \( w(u) > \delta \) if \( u \) is a parent of a leaf. For all other nodes \( u \), we inductively select
\[
w(u) \in \left( \sum_{v \in C(u)} w(v), \sum_{v \in C(u)} w(v) + \frac{1}{k} \right)
\]
in a bottom-up fashion. It is clear that this can be done in such a way that the resulting \( w \) is generic; by construction, \( w \) also satisfies the condition (3).

**Lemma 14.** With \( w \) as previously defined, we have
\[
w_k^\delta(G) \geq w(T(G)) - 1
\]
and
\[
w_k^\delta(G \setminus \{h\}) < w(T(G)) - 1
\]
for every interior node \( h \in G \); here \( T(G) \) is the set of top nodes in \( G \), the \( \delta/(k+1) \) children of the root.

In particular, \( w^\delta(G \setminus \{h\}) < w^\delta(G) \) for all interior nodes \( h \in G \), so \( h \) is an extreme constraint in \( G \) w.r.t. the \( k \)-optimized LP-type problem \( T_k \). With Lemma 7 we obtain the following

**Corollary 2.** With the tree structure and weights as above, the combinatorial dimension of the \( k \)-optimized LP-type problem \( T_k \) is \( \Theta(\delta^{k+1}) \).
Proof. (Lemma 14.) For the lower bound on \( w_k^\delta(G) \), we successively remove the “best \( k \)” constraints from \( G \) (in increasing order of distance to the root) and look at the values \( w^\delta(G) = w^\delta(G_0), w^\delta(G_1), \ldots, w^\delta(G_k) \), where \( G_i \) is the set after removing \( i \) constraints. The value \( w^\delta(G_k) \) determines \( w_k^\delta(G) \). We first observe that \( B(G_i) = T(G_i) \) for all \( i \), because the degree of the tree allows us to increase the number of top nodes by at most \( \delta/(k+1) \) per removal step, so we will never have more than \( \delta \) top nodes in the process.

Now, if the \( i \)-th removed constraint is a top node in \( G_{i-1} \), none of its children have been removed yet, so all of them become top nodes in \( G_i \), and by the previous observation along with (5) we obtain

\[
w^\delta(G_i) \geq w^\delta(G_{i-1}) - \frac{1}{k}.
\]

If the \( i \)-th removed constraint is not a top node in \( G_{i-1} \), we get \( T(G_i) = T(G_{i-1}) \) and hence

\[
w^\delta(G_i) = w^\delta(G_{i-1}).
\]

After \( k \) removal steps, we therefore have

\[
w_k^\delta(G) = w^\delta(G_k) \geq w^\delta(G_0) - 1 = w(T(G)) - 1.
\]

For the upper bound, we start from \( G_0 = G \setminus \{h\} \) for some interior node \( h \) and choose any path from a leaf to the root that contains node \( h \). This path has exactly \( k \) interior nodes in \( G \setminus \{h\} \). Removing all of them in turn, and in increasing order of distance to the root, monotonicity of \( w^\delta \) yields

\[
w^\delta(G_{k-1}) \leq w^\delta(G_0) \leq w^\delta(G) = w(T(G)),
\]

where \( G_{k-1} \) is the set obtained before the last removal step, the one that removes a parent \( p \) of a leaf and yields \( G_k \). In that step, all children of \( p \) become top nodes. By our choice of \( w \), and again observing that throughout the process, \( B(G_1) = T(G_1) \), we have

\[
w^\delta(G_k) = w^\delta(G_{k-1}) - w(p) + \sum_{v \in C(p)} w(v)
\]

\[
< w^\delta(G_{k-1}) - \delta + \frac{\delta}{k+1} < w^\delta(G_{k-1}) - 1.
\]

Together with the previous inequality, this yields

\[
w_k^\delta(G \setminus \{h\}) \leq w^\delta(G_k) < w(T(G)) - 1,
\]

since \( G_k \) is a \( k \)-candidate of \( G \setminus \{h\} \).

\[
\]

7.3 A Geometric Lower Bound

The top nodes LP-type problem is an artificial one for which we do not claim any geometric realizability; however, it directly generalizes the following geometric lower bound construction for the case \( k = 1 \) (removal of one constraint).
Theorem 6. Let \( d \geq 2 \), and let \( H \subseteq \mathbb{R}^{2d} \) consist of the following points (where \( \approx \) indicates a sufficiently small perturbation to achieve nondegeneracy); \( e_i \) is the \( i \)-th unit vector.

\[
\begin{align*}
p_i &\approx e_i, \quad i = 1, 2, \ldots, d, \\
p_{ij} &\approx \frac{9}{10}e_i + \frac{1}{10}e_j, \quad i = 1, 2, \ldots, d, \quad j = d + 1, d + 2, \ldots, 2d.
\end{align*}
\]

Let \( (H, \Omega, w') \) denote the strongly nondegenerate refinement of the LP-type problem associated with the SEB problem over \( H \), as introduced in Section 3.2. Concretely, \( w'(G) = (w(G), V_{SEB}(G)) \) (pairs are compared lexicographically), where \( w(G) \) is the radius of the smallest enclosing ball of \( G \subseteq H \), and \( V_{SEB}(G) \) is the set of constraints violated by \( G \) (points outside of the smallest enclosing ball). Then the following two statements hold.

(i) The LP-type problem \( (H, \Omega, w') \) has combinatorial dimension at most \( 2d + 1 \).

(ii) The 1-optimized LP-type problem \( (H, \Omega, w'_1) \) has combinatorial dimension at least \( d^2 \).

Part (i) simply follows from \( H \subseteq \mathbb{R}^{2d} \). Before we prove part (ii), let us discuss the intuition. Up to a small perturbation, the set \( H \) consists of the first \( d \) unit vectors, plus for each of them a group of \( d \) points “nearby”, but closer to the centroid \( c \) of \( p_1, p_2, \ldots, p_d \). This ensures that the smallest enclosing ball of \( H \) has center \( c \) and exactly \( p_1, p_2, \ldots, p_d \) on its boundary. After removing \( p_i \), however, the smallest enclosing ball retracts, and \( p_i \) is replaced as a boundary point by all the \( p_{ij} \) from its nearby group (here we need the extra dimensions). We can formalize this as follows and omit the elementary proof.

Lemma 15. Let \( (H, \Omega, w') \) be as above.

(i) Exactly the points \( p_i, i = 1, 2, \ldots, d, \) are extreme in \( H \) w.r.t. \( w' \).

(ii) For \( i = 1, 2, \ldots, d \), exactly the points \( p_{\ell}, \ell \neq i \) and \( p_{ij}, j = d + 1, d + 2, \ldots, 2d \), are extreme in \( H \setminus \{p_i\} \) w.r.t. \( w' \).

Proof. (Theorem 6). We will show that Lemma 15 implies that all points \( p_{ij} \) are extreme in \( H \) w.r.t. \( w'_1 \). This means that the combinatorial dimension of the 1-optimized LP-type problem \( (H, \Omega, w'_1) \) is at least \( d^2 \), as claimed.

Let \( C_{ij} = H \setminus \{p_{ij}, h\} \) be the unique 1-witness of \( H \setminus \{p_{ij}\} \). We first observe that \( h \) must be one of the \( p_\ell \)’s, since otherwise, \( C_{ij} \) would still have the same smallest enclosing ball as \( H \), a contradiction to \( C_{ij} \) being the result of “best 1” removal from \( H \setminus \{p_{ij}\} \).

It actually follows that \( h = p_i \). To see this, suppose \( h = p_\ell, \ell \neq i \). Since \( p_{ij} \) is not extreme in \( H \setminus \{p_\ell\} \), we get

\[
w'(H \setminus \{p_\ell\}) = w'(H \setminus \{p_{ij}, p_\ell\}) = w'(C_{ij}).
\]

But then \( C_{ij} \) cannot be the 1-witness of \( H \setminus \{p_{ij}\} \), since there is a better 1-candidate \( H \setminus \{p_{ij}, p_i\} \): we have (by Lemma 15(ii) and the slight perturbation) that

\[
w'(H \setminus \{p_{ij}, p_i\}) < w'(H \setminus \{p_i\}) \approx w'(H \setminus \{p_\ell\}) = w'(C_{ij}).
\]
We have now shown that $C_{ij} = H \setminus \{p_{ij}, p_i\}$, and using Lemma 15(ii) again, we have
\[ w'(C_{ij} \cup \{p_{ij}\}) = w'(H \setminus \{p_i\}) > w'(H \setminus \{p_{ij}, p_i\}) = w'(C_{ij}). \]

Via Lemma 6(ii), this inequality shows that $p_{ij}$ is extreme in $H$ w.r.t. $w'_1$. \qed

8 Discussion and Open Problems

If an LP-type problem $(H, \Omega, w)$ with $|H| = n$ constraints has combinatorial dimension $\delta$, the Sampling Lemma 1 together with Lemma 7 (every set has at most $\delta$ extreme constraints) implies that
\[ v_r \leq \delta \frac{n - r}{r + 1}. \]

(6)

Here, $v_r$ is the expected number of constraints violated by a random sample $R \subseteq H$ of size $|R| = r$.

This, however, is only a worst case bound. The true value of $v_r$ as given by the Sampling Lemma is
\[ v_r = x_{r+1} \frac{n - r}{r + 1}, \]

where $x_{r+1}$ is the average number of extreme constraints in a random sample of size $r + 1$. We know that $x_{r+1} = \delta$ if $r + 1 = n$ and the basis of $H$ happens to be among the ones of maximal size $\delta$. In other cases, $x_{r+1}$ might be much smaller than the combinatorial dimension $\delta$. In particular, in Clarkson’s algorithms [4, 7] and also in chance-constrained optimization [1, 2], the Sampling Lemma is used with sublinear values of $r$, and this is the whole point of the sampling approach.

Our lower bound result from the previous section has therefore no practical value in the applications: We cannot use it to construct an LP-type problem where sampling $r$ constraints at random and then removing the “best $k$” yields
\[ v_r = \Omega(\delta^{k+1} \frac{n - r}{r + 1}) \]

for relevant values of $r$ as a function of $n$.

On the positive side, we show that if better bounds on $v_r$ exist for relevant cases, they cannot be obtained with the help of just the combinatorial dimension. This is an important insight, since in all theoretical work that we are aware of, it is precisely the inequality (6) that is used to bound the expected number of violated constraints.

Our work therefore leads to the obvious open problem of finding bounds on $x_{r+1}$ that improve over $\delta$ in relevant cases.

In practice, it might be prohibitively expensive to remove the “best $k$” constraints, so it is natural to ask whether other removal strategies lead to similar theoretical results. In the abstract setting of LP-type problems, our techniques do not seem to leave much room for this; after all, we crucially need the property that all removed constraints are violated.
by the $k$-optimized solution, and to prove this in Lemma 3, we needed a $k$-witness to have the smallest possible $w$-value among all $k$-candidates.

Another question is whether the requirement of strong nondegeneracy underlying our results can be relaxed. As pointed out in Section 5, it suffices to assume “normal” nondegeneracy, and in many geometric situations, this can be achieved through a (symbolic) perturbation of the input. In the abstract setting of LP-type problems, however, nondegeneracy is a more restrictive assumption: As shown by Matoušek and Škovroň [11], degeneracies may be removable only at the price of a substantial increase in the problem dimension.

Finally, the geometric lower bound in Section 7.3 for the case $k = 1$ suggests that also for $k > 1$, our artificial lower bound example (the top nodes LP-type problem) might be replaceable by a geometric lower bound example. We have not been able to work this out so far.

8.1 Acknowledgments

I thank Georg Schildbach and Lorenzo Fagiano for bringing sampling with removal to my attention, and for many fruitful discussions. I also thank the reviewers as well as May Szedlak for many useful comments that helped to improve the presentation.

References

[1] M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. SIAM J. Optim., 19:1211–1230, 2008.

[2] M. C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. J. Optim. Theory Appl., 148:257–280, 2011.

[3] K. L. Clarkson. A bound on local minima of arrangements that implies the upper bound theorem. Discrete Comput. Geom., 10:427–233, 1993.

[4] K. L. Clarkson. Las Vegas algorithms for linear and integer programming. J. ACM, 42:488–499, 1995.

[5] L. Fagiano and G. Schildbach. Sampling lemma with a-posteriori sample removal. Manuscript, 2013.

[6] B. Gärtner. Sampling with removal in LP-type problems. In Proc. 13th Annu. ACM Sympos. Comput. Geom., pages 511–518, 2014.

[7] B. Gärtner and E. Welzl. Linear programming – randomization and abstract frameworks. In Proc. 13th Sympos. Theoret. Aspects Comput. Sci., volume 1046 of Lecture Notes Comput. Sci., pages 669–687. Springer-Verlag, 1996.
[8] B. Gärtner and E. Welzl. A simple sampling lemma - analysis and applications in geometric optimization. *Discrete Comput. Geom.*, 25(4):569–590, 2001.

[9] J. Matoušek. On geometric optimization with few violated constraints. *Discrete Comput. Geom.*, 14:365–384, 1995.

[10] J. Matoušek, M. Sharir, and E. Welzl. A subexponential bound for linear programming. In *Proc. 8th Annu. ACM Sympos. Comput. Geom.*, pages 1–8, 1992.

[11] J. Matoušek and P. Škovroň. Removing degeneracy may require unbounded dimension increase. *Electronic Notes in Discrete Mathematics*, 29(0):107–113, 2007.

[12] M. Sharir and E. Welzl. A combinatorial bound for linear programming and related problems. In *Proc. 9th Sympos. Theoret. Aspects Comput. Sci.*, volume 577 of *Lecture Notes Comput. Sci.*, pages 569–579. Springer-Verlag, 1992.

[13] E. Welzl. Smallest enclosing disks (balls and ellipsoids). In H. Maurer, editor, *New Results and New Trends in Computer Science*, volume 555 of *Lecture Notes Comput. Sci.*, pages 359–370. Springer-Verlag, 1991.