A Robust Robust Optimization Result *

Martina Gancarova and Michael J. Todd †

November 20, 2018

Abstract

We study the loss in objective value when an inaccurate objective is optimized instead of the true one, and show that "on average" this loss is very small, for an arbitrary compact feasible region.

1 Introduction

This paper is concerned with the loss in objective value incurred when an inaccurate objective, because of either uncertainty or misspecification, is optimized instead of the true one.

Consider the following model case. Instead of the true objective $w^T x$, the nominal objective $v^T x$ is maximized over the unit ball in $\mathbf{R}^n$, where w and v are unit vectors making an angle α. (Throughout the paper, we assume $0 < \alpha < \pi/2$.) Then the computed optimal solution is $x = v$, attaining a true objective value of $\cos\alpha$. Since the true optimal value is 1, the loss is $1 - \cos\alpha$, but to make the measure scale-invariant, we divide by the range of the true objective over the feasible region, which is 2 (from -1 to $+1$). The scaled loss is thus $(1 - \cos\alpha)/2$: see Figure 1. Our main result claims that this formula for the scaled loss holds "on average" for any compact feasible region. Since this result on the robustness of the optimal value to misspecification of the objective holds for any feasible region, we call it a robust robust optimization result.

Robust optimization has been much studied over the last fifteen years: see, e.g., [2, 3, 4, 5, 6]. Usually there is uncertainty in the constraints as well as the objective, and the goal is to find a decision vector that is feasible regardless of the realization of the constraints and that achieves a guaranteed performance regardless of the realization of the objective. Typically this leads to an optimization problem that is harder than the deterministic version of the problem. Our concerns are appropriate when the decision maker is oblivious to the error in the objective and does not protect against a possible misspecification.

In Section 2 we define our setting and give a worst-case bound on the scaled loss for a class of feasible regions. Section 3 describes two probability distributions for the true and nominal objectives and obtains our probabilistic result; we also explain why "on average" is in quotes above. Finally, in Section 4 we discuss the result and outline two applications.

*Supported in part by NSF through grant DMS-0513337 and ONR through grant N00014-08-1-0036.
†School of Operations Research and Information Engineering, Cornell University, Ithaca, New York 14853, USA

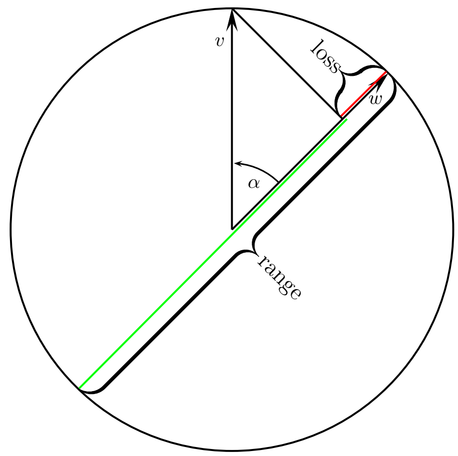

Figure 1: Scaled loss: model case. The loss is the red segment; the range the sum of the red and green segments.

2 Definitions and Worst-Case Results

Let $C \subset \mathbf{R}^n$ be compact and nonempty. Since we are interested in maximizing linear functions over C, we could assume without loss of generality that C is convex, by replacing it if necessary by its convex hull. However, in Section 4 we treat optimization over a nonlinear transformation of a compact set and over a set of binary vectors, and so we prefer not to restrict C further.

Definition 1. *For $v \in \mathbf{R}^n$, we define*

$$\begin{aligned}\max(v) &:= \max\{v^T x : x \in C\},\\ \min(v) &:= \min\{v^T x : x \in C\}, \text{ and}\\ \operatorname{range}(v) &:= \max(v) - \min(v).\end{aligned}$$

Now consider two objectives, the *true* objective $w^T x$ and the *nominal* objective $v^T x$. If we maximize $v^T x$ over C, the optimal solution set is $\{x \in C : v^T x = \max(v)\}$, and if this is not a singleton, we might be unlucky and choose the worst x as far as the true objective is concerned. Hence we make

Definition 2. *For $v, w \in \mathbf{R}^n$, define*

$$\begin{aligned}\operatorname{loss}(v,w) &:= \max(w) - \min\{w^T x : x \in C, v^T x = \max(v)\} \text{ and}\\ \operatorname{scaled\_loss}(v,w) &:= \frac{\operatorname{loss}(v,w)}{\operatorname{range}(w)}.\end{aligned}$$

Note that the scaled loss is invariant to translations or dilations of C, and even to rotations if v and w are correspondingly rotated.

As we have seen, if $C = B^n := \{x \in \mathbf{R}^n : \|x\| \le 1\}$ (all norms are Euclidean), $v = (0; 1; 0; \ldots; 0)$, and $w = (\sin\alpha; \cos\alpha; 0; \ldots; 0)$, then $\operatorname{loss}(v,w) = 1 - \cos\alpha$ and $\operatorname{scaled\_loss}(v,w) = (1-\cos\alpha)/2$. On the other hand, if C is the convex hull of $(-1; 0; \ldots; 0)$ and $(+1; 0; \ldots; 0)$ and v and w are as above, then $\operatorname{loss}(v,w) = \operatorname{range}(w) = 2\sin\alpha$ and the scaled loss is 1, as bad as it can be.

Note that in this example, the optimal solution set for $v^T x$ is all of C, and in accordance with the definition above, we choose the worst of these optimal solutions with respect to the true objective, namely $(-1; 0; \ldots; 0)$, in evaluating the loss.

Also, for $n = 2$, we will usually view the nominal objective v as pointing vertically up as in this example. Observe that there is a subtle bias in this viewpoint. While the decision maker only sees v, and therefore thinks of v as fixed and w (if she thinks of it at all) as a perturbation of v, a more appropriate perspective would regard the true objective w as being generated in some suitable way, and then v arising as a perturbation of w.

In the rest of this section, we obtain a worst-case bound on the scaled loss when C is restricted to avoid the situation above.

Theorem 1. *Assume that C is contained in B^n and contains rB^n for some $0 < r < 1$. Let v and w be two nonzero vectors making an angle α, where $\sin\alpha \le r \le \cos\alpha$. Then, with $\rho := \sqrt{1-r^2}$, we have*

$$\text{scaled\_loss}(v, w) \le \frac{2\rho\sin\alpha}{r(1+\cos\alpha) + \rho\sin\alpha},$$

and this bound is tight.

Proof. First we show that the right-hand side above can be attained. Let $n = 2$ and choose C to be the convex hull of $(-\rho; r)$, $(\rho; r)$, and rB^2. Let $v = (0; 1)$ and $w = (\sin\alpha; \cos\alpha)$. Then the set of optimal solutions for $v^T x$ is the convex hull of $(-\rho; r)$ and $(\rho; r)$, with the former being worst for $w^T x$. With our assumption that $r \le \cos\alpha$, the optimal solution for $w^T x$ is $(\rho; r)$, while $r \ge \sin\alpha$ implies that $w^T x$ is minimized at $-rw$. Hence the loss is $2\rho\sin\alpha$ and the range $r(1+\cos\alpha) + \rho\sin\alpha$, giving the scaled loss as indicated. See Figure 2.

Now we need to prove the bound. Given any C, v, and w, we can project C into the plane spanned by v and w, and the projected C will lie in B^2 and contain rB^2. Hence we can assume that $n = 2$. By rotating if necessary, we can assume that v and w are as above (note that the scale of these vectors is immaterial). Let $s := \max\{v^T x : x \in C\} \ge r \ge \sin\alpha$, and let $\sigma := \sqrt{1-s^2}$. Then $\min\{w^T x : x \in C, v^T x = \max(v)\} \ge -\sigma\sin\alpha + s\cos\alpha \ge 0$ and $\min(w) \le -r$, and so

$$\text{scaled\_loss}(v, w) \le \frac{\max(w) + \sigma\sin\alpha - s\cos\alpha}{\max(w) + r};$$

note that the right-hand side is monotonically increasing in $\max(w)$, so substituting an upper bound for the latter provides a valid upper bound on the scaled loss.

If $s \le \cos\alpha$, then $\max(w) \le \sigma\sin\alpha + s\cos\alpha$, and we deduce

$$\text{scaled\_loss}(v, w) \le \frac{2\sigma\sin\alpha}{\sigma\sin\alpha + s\cos\alpha + r} \le \frac{2\rho\sin\alpha}{\rho\sin\alpha + r\cos\alpha + r},$$

as desired.

On the other hand, if $s > \cos\alpha$ so that $\sigma < \sin\alpha$, then $\max(w) \le 1$ and so

$$\text{scaled\_loss}(v, w) \le \frac{1 + \sigma\sin\alpha - s\cos\alpha}{1 + r} \le \frac{1 + \sin^2\alpha - \cos^2\alpha}{1 + r} = \frac{2\sin^2\alpha}{1 + r} \le \frac{2\rho\sin\alpha}{r\cos\alpha + \rho\sin\alpha + r},$$

since $\sin\alpha \le \rho$ and $1 \ge r\cos\alpha + \rho\sin\alpha$. Hence the bound is established in either case.
□

Since $\rho < \cos\alpha$, the right-hand side above is at most $\sin\alpha / r$, and it approaches this value for small r and very small α. This bound is of order α / r, and hence much larger than $(1 - \cos\alpha)/2 \approx \alpha^2/4$, the value in the model case.

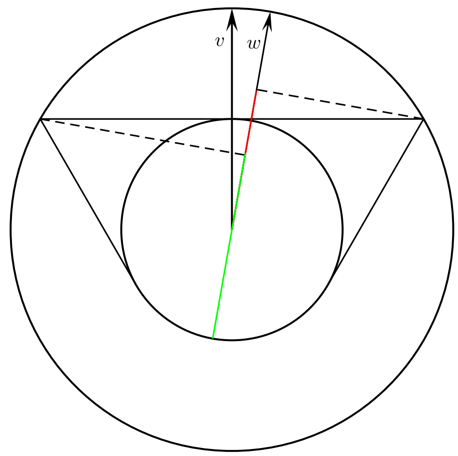

Figure 2: Scaled loss: worst case. The feasible region is the convex hull of the horizontal line segment and the smaller circle. Again, the loss is the red segment; the range the sum of the red and green segments.

3 Probabilistic Analysis

Now we examine how the scaled loss behaves when v and w are generated randomly. We examine two different probability distributions. We call $v \in \mathbf{R}^n$ a *standard Gaussian vector* if its components are independent standard Gaussian random variables, or equivalently if $v \sim N(0, I)$.

Definition 3. *We say (w, v) is generated according to Probability Distribution 1 if w and u are independent standard Gaussian vectors in $\mathbf{R}^n$, and $v := w \cos\alpha + u \sin\alpha$. Expectations with respect to this distribution are indicated by E_1.*

We say a random variable ξ depending on n *concentrates* around a positive constant β if, for every positive δ, the probability that ξ lies between $(1-\delta)\beta$ and $(1+\delta)\beta$ converges to 1 as $n \to \infty$.

Proposition 1. *The angle between v and w generated according to Probability Distribution 1 concentrates around α as $n \to \infty$.*

Proof. Because w and u are standard Gaussian random vectors, so is v, and both $w^T w$ and $v^T v$ are chi-squared random variables with n degrees of freedom, both concentrated around their means, n. Also, $v^T w = \cos\alpha\, w^T w + \sin\alpha\, u^T w$, and since $w^T w$ is concentrated around n and $u^T w$ has mean zero and variance $O(n)$, this is concentrated around $\cos\alpha\, n$. Hence, using a union bound, we find that with probability approaching 1, $v^T w/(v^T v\, w^T w)^{1/2}$ lies between $(1-\epsilon)\cos\alpha/(1+\epsilon)$ and $(1+\epsilon)\cos\alpha/(1-\epsilon)$ for any positive ϵ. This implies the result. $\square$

We now define our second model:

Definition 4. *We say (w, v) is generated by Probability Distribution 2 if $\bar{w}$ and $\bar{u}$ are independent standard Gaussian vectors in $\mathbf{R}^n$, $\hat{u} = (I - \bar{w}\bar{w}^T/\bar{w}^T\bar{w})\bar{u}$, $w = \bar{w}/\|\bar{w}\|$, $u = \hat{u}/\|\hat{u}\|$, and $v = w\cos\alpha + u \sin\alpha$. Expectations with respect to this distribution are indicated by E_2.*

Note that all the vectors are well-defined with probability one, with w and u orthogonal vectors having unit norm, so that w and v are unit vectors making an angle α with probability one.

In both distributions, w is generated according to some distribution, and v is generated as a perturbation of w. This fits with our interpretation of w as the true objective and v as a nearby inaccurate objective. However, as we now show, we can alternatively regard v as being generated first and then w as a perturbation of v. This is very useful in our analysis.

Proposition 2. *In both Probability Distribution 1 and Probability Distribution 2,* $(v, w) \sim (w, v)$.

Proof. Consider first Probability Distribution 1. Since the matrix

$$\begin{bmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{bmatrix}$$

is orthogonal, $(v, z) := (w\cos\alpha + u\sin\alpha, w\sin\alpha - u\cos\alpha) \sim (w, u)$. Then, since (v, w) is derived from the first pair exactly as (w, v) is derived from the second, we see that $(v, w) \sim (w, v)$.

Next assume (w, v) is generated according to Probability Distribution 2. Then w lies on the unit $(n-1)$-dimensional sphere $S^{n-1} := \{x \in \mathbf{R}^n : \|x\| = 1\}$, and $v = w\cos\alpha + u\sin\alpha$, where u lies on the unit $(n-2)$-dimensional sphere $S_w^{n-2} := \{x \in \mathbf{R}^n : w^T x = 0, \|x\| = 1\}$.

Let Q be any orthogonal matrix. Since $(Q\bar{w}, Q\bar{u}) \sim (\bar{w}, \bar{u})$, and $Q\bar{w}/\|Q\bar{w}\| = Qw$, we see that the distribution of Qw coincides with that of w, so that w is uniformly distributed on S^{n-1}. Also, $(I - Q\bar{w}\bar{w}^T Q^T/(Q\bar{w})^T(Q\bar{w}))Q\bar{u} = Q(I - \bar{w}\bar{w}^T/\bar{w}^T\bar{w})\bar{u} = Q\hat{u}$, so that $(Q\bar{w}, Q\bar{u})$ gives rise to Qw and Qu and hence $Qv = (Qw)\cos\alpha + (Qu)\sin\alpha)$. It follows that v is also distributed uniformly on S^{n-1}.

Moreover, even if Q is taken to be an orthogonal matrix that fixes w, so that Q depends on w, we still have $Q\bar{w} = \bar{w}$ a standard Gaussian vector in $\mathbf{R}^n$ and $Q\bar{u}$ a standard Gaussian vector independent of w. Hence, proceeding as above, we see that $(w = Q\bar{w}, Q\bar{u})$ gives rise to $Q\hat{u}$ and hence Qu, so since $(\bar{w}, Q\bar{u}) \sim (\bar{w}, \bar{u})$, we find $Qu \sim u$, from which, conditioned on w, u is uniformly distributed on S_w^{n-2}.

Finally, we show that $(v, z) := (w\cos\alpha + u\sin\alpha, w\sin\alpha - u\cos\alpha) \sim (w, u)$. Since $w = v\cos\alpha + z\sin\alpha$, this will show that $(v, w) \sim (w, v)$ as desired. We have already shown that v, the first member of the pair, is distributed uniformly on S^{n-1}. We now consider all pairs $(\bar{w}, \bar{u})$ that give rise to a given v. If Q is an orthogonal matrix that fixes v, then $(Q\bar{w}, Q\bar{u})$ also gives rise to the same v. Since Q is orthogonal, the distribution of $(\bar{w}, \bar{u})$, conditional on this fixed v, is invariant under pre-multiplication of each vector by such a Q. As we have seen, under this transformation w is transformed to Qw and u to Qu, and hence $z := w\sin\alpha - u\cos\alpha$ is transformed to Qz. It follows that z, which has unit norm and is orthogonal to v, is uniformly distributed on S_v^{n-2}. This concludes the proof. □

We are now ready to analyze the behavior of the scaled loss "on average" for our two models. First we investigate the range function:

Lemma 1. *For* $i = 1, 2$, *we have*

$$E_i \max(v) = E_i \max(w), \quad E_i \operatorname{range}(w) = 2E_i \max(w).$$

Proof. The first equation follows from Proposition 2 above, since v and w have the same distribution. For the second equation, note that

$$E_i \operatorname{range}(w) = E_i \max(w) - E_i \min(w) = E_i \max(w) + E_i \max(-w),$$

and that $E_i \max(-w) = E_i \max(w)$ since under both probabilistic models, w has a symmetric distribution. □

Next we examine the loss:

Lemma 2. *For $i = 1, 2$, we have*

$$E_i \operatorname{loss}(v, w) = (1 - \cos\alpha) E_i \max(w).$$

Proof. First note that, since C is compact, the convex function $\max(v)$ is finite everywhere, and hence is differentiable almost everywhere, with respect to Lebesgue measure and hence with respect to Probability Distribution 1. But $\max(v)$ is differentiable at v exactly when the maximum of $v^T x$ over C is attained at a single x, which we denote by x_v. Since this property is invariant under positive scalings of v, we see that it holds also for almost all v under Probability Distribution 2 also.

Hence with probability one,

$$\operatorname{loss}(v, w) = \max(w) - w^T x_v = \max(w) - (v \cos\alpha + z \sin\alpha)^T x_v,$$

where, as in the proof of Proposition 2, we let $z := w \sin\alpha - u \cos\alpha$. Now in either model, z has a symmetric distribution conditional on v, and since with probability one x_v depends only on v, $z^T x_v$ has mean zero. Hence

$$E_i \operatorname{loss}(v, w) = E_i \max(w) - \cos\alpha\, E_i \max(v) - 0,$$

and the result follows from Lemma 1. $\square$

From Lemmas 1 and 2, we immediately deduce

Theorem 2. *For $i = 1, 2$, we have*

$$\frac{E_i \operatorname{loss}(v, w)}{E_i \operatorname{range}(w)} = \frac{1 - \cos\alpha}{2}.$$

$\square$

Note that we do not have a result on the expected scaled loss, which would be an expectation of the ratio of the loss to the range, but only on the ratio of the expectations, which is why we have put “on average” in quotes above.

4 Discussion and applications

It seems at first that the theorem of the last section would hold under much weaker probabilistic assumptions, merely requiring that w and u have symmetric distributions. Unfortunately, there are two problems with this. First, what we really need is that v and z have symmetric distributions, but putting restrictions on v and z conflicts with the natural interpretation that the true objective w should be generated first, and then v as a perturbation of w. Second, it is crucial that $\max(v)$ and $\max(w)$ have the same expectation, and this appears hard to ensure under weaker assumptions: the fact that (v, w) and (w, v) have the same distribution under our two models is key in our development.

One way in which the result can be generalized is in allowing a random choice of α. Our two models yield vectors v and w making an angle that either concentrates around α or is exactly α. Instead, we can consider probability distributions on the triple (α, w, v) as follows: first α is generated according to an arbitrary distribution supported on $(0, \pi/2)$; then, conditional on α, w and v are generated according to Probability Distribution 1 or 2. It is easy to see that all our arguments can be extended by first conditioning on α, and the expected loss divided by the expected range will be $(1 - E(\cos\alpha))/2$, where the expectation is taken with respect to the distribution on α.

Another generalization allows very general distributions, but changes the way the objective vector is perturbed. Let f_j be a symmetric probability density on $\mathbf{R}$ for $j = 1, \ldots, n$. For each j, draw w_j and u_j independently from f_j, and then let v_j be w_j with probability $\cos\alpha$ and u_j with probability $1 - \cos\alpha$, with all the choices independent. Let $t \in \mathbf{R}^n$ be defined by $t_j = +1$ if $v_j = w_j$, $t_j = -1$ if $v_j = u_j$, so each t_j is $+1$ with probability $\cos\alpha$ and -1 with probability $1 - \cos\alpha$. Then it is clear how (w, v) arises as a function of the triple (w, u, t). But if we define $z := w + u - v$, then in each component, z agrees with w (u) exactly when v agrees with u (w). It follows that (v, z, t) has the same distribution as (w, u, t), and (v, w) arises from (v, z, t) as does (w, v) from (w, u, t). Hence (v, w) and (w, v) have the same distribution. Moreover, the arguments of the previous section can be duplicated, and again lead to the result that the expected loss divided by the expected range is exactly $(1 - \cos\alpha)/2$. Under mild conditions on the f_j's, the angle between v and w concentrates around α. Note that, in this model, a small fraction of the components are changed a possibly large amount, while in the previous models, each component is changed a small amount.

We argued in the introduction that the scaled loss provided a good measure of how much is lost in objective value when implementing the optimal solution for a misspecified objective. However, the result of the previous section is concerned with the ratio of the expectations of the loss and the range, rather than the more meaningful expectation of the ratio. We therefore conducted some experiments with two NETLIB [9] problems, AGG and BOEING1, to see how much the results differ. The first has 489 rows and 163 columns, the second 351 rows and 384 columns. Figures 3 and 4, which give graphs of the expectation of the ratio and of the ratio of the expectations as functions of the angle α in degrees, show that our results should be applicable to the more meaningful measure also. In both figures, each data point is obtained using at least 10,000 pairs (v, w) generated from Probability Distribution 1.

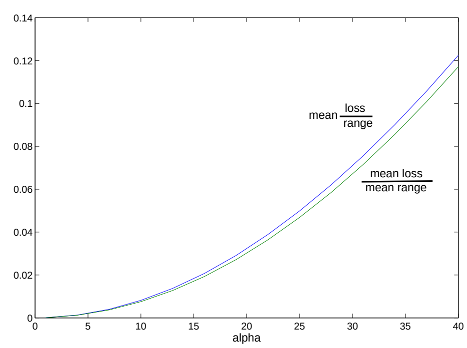

Figure 3: Ratio of expectations versus expectation of ratios for AGG.

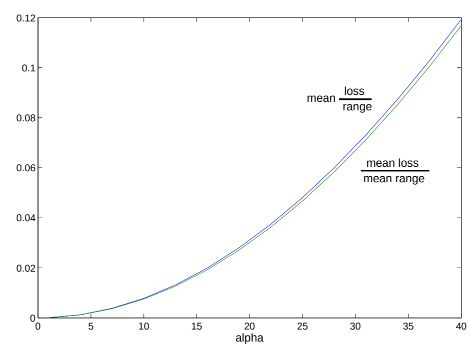

Figure 4: Ratio of expectations versus expectation of ratios for BOEING1.

Our result is limited to clarifying what happens when linear objective functions are perturbed. In general, little can be said for nonlinear functions, partially because it is not clear how random nonlinear objective functions and their perturbations should be defined. However, our analysis can be applied to one case where objective functions are nonlinear. Suppose there are several continuous objective functions, $f_i(x)$ for $i = 1, \ldots, k$. The decision maker is interested in high values of all of these objectives, so we are in the realm of multi-criteria optimization; see, for instance, [7, 8]. Often, a linear combination $\sum_i w_i f_i(x)$ of the objectives is maximized, and indeed, any such optimal solution for a positive w is an efficient or Pareto-optimal solution. (The converse is true when all f_i's are concave, and in this case the optimization problem is convex, but this restriction is not needed for our discussion.) However, since the different function f_i may be hard to compare, it is difficult to decide on appropriate weights w. Our theorem indicates in some sense that the choice may not matter too much.

Let $Y := \{y = (f_1(x); \ldots, f_k(x)) : x \in X\}$. Since X is nonempty and compact, so is Y, and trivially $\max\{\sum_i w_i f_i(x) : x \in X\} = \max\{w^T y : y \in Y\}$. Theorem 2 shows that the latter is insensitive in some precise sense to the specification of the objective w, and this translates directly into the insensitivity of the original problem's optimal value to the specification of the weights. Of course, there is a large caveat here: the result requires v and w to be randomly chosen from symmetric distributions, and hence their components are as likely to be negative as positive, while in the multi-criteria setting, the weights are always positive. Nevertheless, we believe our theorem gives some credence to the hope that incorrect choices of weights should not hurt much.

Our second application is to the complexity of combinatorial optimization problems. Beier and Vöcking [1] and Röglin and Teng [10] have uncovered a fascinating connection between binary optimization problems that can be solved in randomized pseudo-polynomial time, that is, in randomized polynomial time if the data are encoded in unary, and smoothed complexity. In particular, Röglin and Teng show that such a problem can be solved in expected time polynomial in the input size and $1/\sigma$, where the adversarily chosen objective function coefficients are perturbed by independent Gaussians with mean zero and variance σ^2 (there are slight technical subleties; see Sections 2 and 6 of [10]). This is normally interpreted as saying that arbitrarily close to any potentially hard instance there are polynomially solvable instances. This gives support to the belief that one would be unlucky to choose a bad instance, in a rather precise and strong sense. Our result provides another avenue to solving such problems. One can *explicitly* make a small random perturbation of the objective function coefficients, thereby obtaining a problem with provably expected polynomial-time complexity. (Note that $w + z$, where z has independent zero-mean Gaussian components with standard deviation σ, is proportional to $w \cos\alpha + u \sin\alpha$, where u is a Gaussian random vector and $\alpha := \arctan\sigma$.) Solving this perturbed problem gives a feasible solution to the original problem, and Theorem 2 gives credence to the hope that this solution will be close to optimal for the true objective function. (For some models of generating w, renormalizing might give a distribution on α, rather than a fixed value, but the extension mentioned at the beginning of this section allows for this possibility.) Of course, our result only proves this "on average," so one would be unlucky to have an objective function where the loss is large, in a certain sense. We believe this viewpoint provides further insight into the notion of smoothed complexity, at least when only the objective function is perturbed.

References

[1] R. Beier and B. Vöcking, *Random knapsack in expected polynomial time*, Journal of Computer and System Sciences 69 (2004), pp. 306–329.

[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, eds., *Robust Optimization*, Mathematical Programming 107, Numbers 1-2, 2006.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization.* Princeton University Press, Princeton, NJ, 2009.

[4] A. Ben-Tal and A. Nemirovski, *Robust solutions of uncertain linear programs*, Operations Research Letters 25 (1999), pp. 1–13.

[5] A. Ben-Tal and A. Nemirovski, *Robust solution of Linear Programming problems contaminated with uncertain data*, Mathematical Programming 88 (2000), pp. 411–424.

[6] D. Bertsimas and M. Sim, *Tractable approximations to robust conic optimization problems*, Mathematical Programming 107 (2006), pp. 5–36.

[7] I. Das and J.E. Dennis, *A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems*, Structural Optimization 14 (1997), pp. 63–69.

[8] C. Hillermeier, *Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach.* Birkhauser, Basel, 2001.

[9] NETLIB linear programming problems, available from `http://www.netlib.org/lp/data`.

[10] H. Röglin and S.-H. Teng, *Smoothed analysis of multiobjective optimization*, in: Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science, 2009, pp. 681–690.