Optim Eng (2010) 11: 319–337
DOI 10.1007/s11081-009-9101-7

Techniques for exploring the suboptimal set

Joëlle Skaf · Stephen Boyd

Received: 6 June 2008 / Accepted: 8 December 2009 / Published online: 21 January 2010

Abstract The ϵ-suboptimal set $\mathcal{X}_\epsilon$ for an optimization problem is the set of feasible points with objective value within ϵ of optimal. In this paper we describe some basic techniques for quantitatively characterizing $\mathcal{X}_\epsilon$, for a given value of ϵ, when the original problem is convex, by solving a modest number of related convex optimization problems. We give methods for computing the bounding box of $\mathcal{X}_\epsilon$, estimating its diameter, and forming ellipsoidal approximations.

Quantitative knowledge of $\mathcal{X}_\epsilon$ can be very useful in applications. In a design problem, where the objective function is some cost, large $\mathcal{X}_\epsilon$ is good: It means that there are many designs with nearly minimum cost, and we can use this design freedom to improve a secondary objective. In an estimation problem, where the objective function is some measure of plausibility, large $\mathcal{X}_\epsilon$ is bad: It means that quite different parameter values are almost as plausible as the most plausible parameter value.

Keywords Convex optimization · Suboptimal set · Support function · Ellipsoidal approximations

This material is based upon work supported by the Focus Center Research Program Center for Circuit & System Solutions award #2003-CT-888, by JPL award #I291856, by Army award #W911NF-07-1-0029, by NSF award #ECS-0423905, by NSF award #0529426, by DARPA award #N66001-06-C-2021, by NASA award #NNX07AEIIA, by AFOSR award #FA9550-06-1-0514, and by AFOSR award #FA9550-06-1-0312.

J. Skaf (✉) · S. Boyd
Information Systems Lab., Electrical Engineering Department, Stanford University, Stanford, CA 94305-9510, USA
e-mail: jskaf@stanford.edu

S. Boyd
e-mail: boyd@stanford.edu

1 The ϵ-suboptimal set

We consider the general convex optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C} \end{array} \tag{1}$$

where $x \in \mathbf{R}^n$ is the optimization variable, f is the (convex) objective function, and $\mathcal{C}$ is the (closed convex) feasible set. We assume that $p^\star$, the optimal value of the problem, is finite, and achieved by at least one point. For $\epsilon \geq 0$, we say that x is ϵ-suboptimal if $x \in \mathcal{C}$ and $f(x) \leq p^\star + \epsilon$. We will let $\mathcal{X}_\epsilon$ denote the set of all ϵ-suboptimal points,

$$\mathcal{X}_\epsilon = \{x \in \mathcal{C} \mid f(x) \leq p^\star + \epsilon\}, \tag{2}$$

which we call the ϵ-suboptimal set. When $\epsilon = \eta|p^\star|$, we refer to $\mathcal{X}_\epsilon$ as the $100\eta\%$-suboptimal set. The set of optimal points is $\mathcal{X}_0$; the problem (1) has a unique optimal point if and only if $\mathcal{X}_0$ is a singleton, i.e., $\mathcal{X}_0 = \{x^\star\}$. The ϵ-suboptimal set $\mathcal{X}_\epsilon$ is convex, and satisfies $\mathcal{X}_\delta \supseteq \mathcal{X}_\epsilon$ for $\delta \geq \epsilon$.

At the crudest level, we can distinguish between two cases. If $\mathcal{X}_\epsilon$ is small, we say that the problem (1) has a *strong minimum*. When $\mathcal{X}_\epsilon$ is large, we say that the problem has a *soft minimum* or a *weak minimum*. These are vague labels, since they depend on what is meant by 'small' or 'large', as well what value of ϵ is used. (We will see that the term *sharp minimum*, however, has a formal definition given below.)

1.1 Our goal

We are interested in an approximate but quantitative characterization of $\mathcal{X}_\epsilon$, for some particular value (or values) of ϵ. As examples, we might like to know how big the ϵ-suboptimal set is, or in which directions (or along which coordinates) it extends far, and in which directions it does not. Our focus will be on techniques that can be used in practical problems, and not on the analysis, or any mathematical details.

In this note we show how several useful quantitative measures of $\mathcal{X}_\epsilon$ can be computed, by solving a modest number of convex optimization problems. The bounding box, which gives the range of each variable x_i over $\mathcal{X}_\epsilon$, can be computed exactly; we will also describe methods for finding approximations of the maximum deviation from $x^\star$ and the diameter of $\mathcal{X}_\epsilon$, and ellipsoidal approximations of $\mathcal{X}_\epsilon$.

The techniques we describe rely on basic ideas and methods of convex analysis and optimization (see, e.g., Ben-Tal and Nemirovski 2001; Boyd and Vandenberghe 2004; Nesterov 2003), but we have not seen them described in the context of suboptimal set characterization. Nevertheless, we feel that these techniques are very useful, and should be widely and routinely applied in practical problems.

1.2 Why

We mention here some general situations in which we might be interested in characterizing $\mathcal{X}_\epsilon$. First suppose that the problem (1) is a decision or design problem: The variables x_i are values that we can choose, $\mathcal{C}$ is the set of acceptable designs, and $f(x)$

is some cost associated with the design or decision x. If $\mathcal{X}_\epsilon$ is small (i.e., the problem (1) has a strong minimum), we conclude that there are few designs that are nearly minimum cost; we gain little design freedom by allowing ϵ-suboptimal designs. On the other hand, if $\mathcal{X}_\epsilon$ is large (i.e., the problem (1) has a weak minimum), we conclude that there are many designs that achieve nearly optimum cost. In this case, we may wish to consider optimizing a secondary objective, or taking some other advantage of the large set of design choices available to us, if we accept ϵ-suboptimal cost.

Now suppose that the problem (1) is some kind of inversion or estimation problem: The variables x_i are parameters that we wish to estimate, $\mathcal{C}$ is the set of possible parameter values, and $f(x)$ gives the implausibility of the value x, given some observations and prior information. An optimal point $x^\star$ is a most plausible point, and therefore a natural one to choose as an estimate of the true value of the parameter. In this setting, ϵ-suboptimal points correspond to parameter values that are almost as plausible as the most plausible point. In this case, if $\mathcal{X}_\epsilon$ is small, we conclude that only parameters near $x^\star$ have near maximum plausibility, which suggests we can have high confidence in our estimate $x^\star$. On the other hand, if $\mathcal{X}_\epsilon$ is large, we conclude that parameter values far from $x^\star$ explain the observations almost as well as $x^\star$; thus, we should have low confidence in the estimate $x^\star$. Depending on exactly what f is, we can sometimes interpret $\mathcal{X}_\epsilon$ as a confidence set for the estimate $x^\star$.

1.3 Local analysis

Our interest here is obtaining quantitative information about $\mathcal{X}_\epsilon$, for a particular problem instance and a particular value of ϵ, and not qualitative information about what $\mathcal{X}_\epsilon$ looks like, for example, in the limit as $\epsilon \to 0$. Nevertheless we briefly mention here the local analysis of the suboptimal set in two different cases, to show the range of possibilities. (For more on local convex analysis, see, e.g., Bertsekas et al. (2003), Borwein and Lewis (2000).)

The smooth unconstrained case. Suppose (1) is unconstrained (i.e., $\mathcal{C} = \mathbf{R}^n$), has a single optimal point $x^\star$, and f is twice differentiable, with $\nabla^2 f(x^\star) > 0$ (and, of course, $\nabla f(x^\star) = 0$). In this case $\mathcal{X}_\epsilon$ can be approximated by the ellipsoid

$$\mathcal{E}_\epsilon = \{x \mid (1/2)(x - x^\star)\nabla^2 f(x^\star)(x - x^\star) \leq \epsilon\}, \tag{3}$$

for ϵ small. This ellipsoid is centered at the optimal point $x^\star$, has a fixed shape, determined by $\nabla^2 f(x^\star)$, and scales as $\sqrt{\epsilon}$. From this ellipsoidal approximation we can derive useful (approximations of) quantitative measures of $\mathcal{X}_\epsilon$. For example, the diameter of $\mathcal{X}_\epsilon$ is approximately the diameter of $\mathcal{E}_\epsilon$:

$$\mathbf{diam}(\mathcal{X}_\epsilon) \approx \mathbf{diam}(\mathcal{E}_\epsilon) = (2\epsilon/\lambda_{\min}(\nabla^2 f(x^\star))^{1/2}.$$

In this case, we see that $\mathcal{X}_\epsilon$ is relatively 'large' when ϵ is 'small' (since its diameter grows like $\sqrt{\epsilon}$). When $\lambda_{\min}(\nabla^2 f(x^\star))$ is small, $\mathcal{X}_\epsilon$ can be quite large, so we expect a weak minimum.

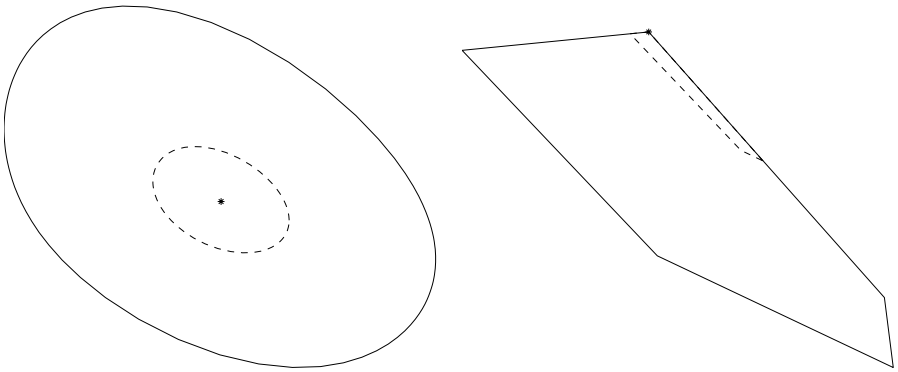

Fig. 1 The boundaries of $\mathcal{X}_{0.01}$ (*dashed line*) and $\mathcal{X}_{0.1}$ (*full line*) for an unconstrained problem (*left*) and a nondifferentiable problem with constraints (*right*). The optimal point $x^\star$ is shown as a *star*

The nondifferentiable (or constrained) case. At the other extreme, suppose that f is piecewise-linear, $\mathcal{C}$ is a polyhedron with nonempty interior, and (1) has a single optimal point $x^\star$. In this case, $\mathcal{X}_\epsilon$ is a polyhedron, and for small ϵ can be exactly represented as

$$\mathcal{X}_\epsilon = \{x \mid A(x - x^\star) \leq 0,\ B(x - x^\star) \leq \epsilon \mathbf{1}\},$$

where A and B correspond to the active constraints defining $\mathcal{C}$, and f, near $x^\star$, and $\mathbf{1}$ denotes the vector with all entries one. (We have $A = 0$ if $x^\star \notin \partial\mathcal{C}$.) In this case $\mathbf{diam}(\mathcal{X}_\epsilon)$ is exactly linear in ϵ, for small ϵ. Moreover, $x^\star$ is generally not well centered in $\mathcal{X}_\epsilon$; indeed, when $x^\star \in \partial\mathcal{C}$, $x^\star$ is on the *boundary* of $\mathcal{X}_\epsilon$.

This case is very different from the smooth unconstrained case considered above. In this case, we expect a sharp minimum, since $\mathbf{diam}(\mathcal{X}_\epsilon)$ grows only linearly with ϵ; moreover, we do not expect $x^\star$ to be well centered in $\mathcal{X}_\epsilon$, in general.

Polyak defines a point $x^\star$ to be a *sharp minimum point* if for some $\alpha > 0$, we have

$$f(x) \geq f(x^\star) + \alpha \|x - x^\star\|_2$$

for all $x \in \mathcal{C}$ (see Polyak 1987, Sect. 5.3). This is the case here.

Figure 1 shows two simple examples illustrating the difference between these two cases. We plot the boundaries of $\mathcal{X}_{0.01}$ and $\mathcal{X}_{0.1}$, for two problems with $x \in \mathbf{R}^2$, one unconstrained smooth problem, and one with piecewise-linear objective and linear constraints. For the smooth problem (shown at left), we see that $\mathcal{X}_\epsilon$ is approximately ellipsoidal and centered around $x^\star$, with $\mathcal{X}_{0.01}$ roughly a factor $\sqrt{10}$ smaller than $\mathcal{X}_{0.1}$. For the nondifferentiable problem (shown at right), we see that $x^\star$ is on the boundary of $\mathcal{X}_\epsilon$, and that $\mathcal{X}_{0.01}$ is roughly a factor 10 smaller than $\mathcal{X}_{0.1}$.

1.4 Support function and boundary points

The *support function* S_ϵ of the ϵ-suboptimal set $\mathcal{X}_\epsilon$ is defined as

$$S_\epsilon(y) = \sup\{y^T x \mid x \in \mathcal{X}_\epsilon\} = \sup\{y^T x \mid x \in C,\ f(x) \leq p^\star + \epsilon\}.$$

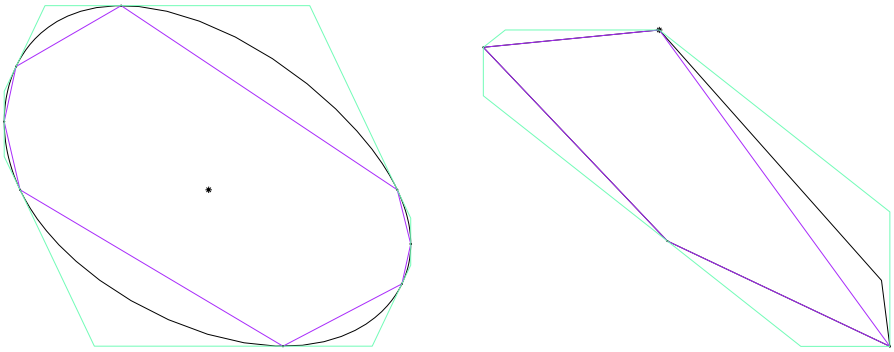

Fig. 2 (Color online) Boundaries of the sets $\mathcal{X}_\epsilon$ (*black*), $\mathcal{P}_{\text{inner}}$ (*purple*), and $\mathcal{P}_{\text{outer}}$ (*cyan*), for an unconstrained problem (*left*) and a nondifferentiable problem with constraints (*right*). The optimal point $x^\star$ is shown as a *star*

It is convex and homogeneous in y, and concave and nondecreasing in ϵ (see Boyd and Vandenberghe 2004, Ex. 3.7.)

Our characterization of $\mathcal{X}_\epsilon$ will rely on our ability to compute $S_\epsilon(y)$, for a given y, by solving the convex problem

$$\begin{array}{ll} \text{maximize} & y^T x \\ \text{subject to} & x \in \mathcal{C}, \quad f(x) \leq p^\star + \epsilon, \end{array} \tag{4}$$

with variable x. Suppose $x^\star(y)$ is a solution of (4). Then $x^\star(y) \in \partial\mathcal{X}_\epsilon$ (i.e., $x^\star(y)$ is on the boundary of $\mathcal{X}_\epsilon$), and we have

$$\mathcal{X}_\epsilon \subseteq \{z \mid y^T(z - x^\star(y)) \leq 0\}.$$

(In fact, $x^\star(y)$ is on the relative boundary of $\mathcal{C}$, if the dimension of $\mathcal{C}$ is less than n.)

Now suppose that we have solved (4) for the (nonzero) values $y^{(1)}, \ldots, y^{(N)}$, with associated optimal points $x^{(1)}, \ldots, x^{(N)}$. From these data we can construct outer and inner polyhedral approximations of $\mathcal{X}_\epsilon$. We have $\mathcal{X}_\epsilon \subseteq \mathcal{P}_{\text{outer}}$, where $\mathcal{P}_{\text{outer}}$ is the polyhedron

$$\mathcal{P}_{\text{outer}} = \{x \mid y^{(i)T}(x - x^{(i)}) \leq 0,\ i = 1, \ldots, N\}.$$

By convexity of $\mathcal{X}_\epsilon$, we also have $\mathcal{X}_\epsilon \supseteq \mathcal{P}_{\text{inner}}$, where $\mathcal{P}_{\text{inner}}$ is the polyhedron

$$\begin{aligned} \mathcal{P}_{\text{inner}} &= \mathbf{conv}\{x^{(1)}, \ldots, x^{(N)}\} \\ &= \{\lambda_1 x^{(1)} + \cdots + \lambda_N x^{(N)} \mid \lambda \geq 0,\ \mathbf{1}^T\lambda = 1\}. \end{aligned}$$

Figure 2 shows $\mathcal{X}_\epsilon$, $\mathcal{P}_{\text{inner}}$ and $\mathcal{P}_{\text{outer}}$ for two problems with $x \in \mathbf{R}^2$, one unconstrained smooth problem, and one with piecewise-linear objective and linear constraints. The sets $\mathcal{P}_{\text{inner}}$ and $\mathcal{P}_{\text{outer}}$ are generated from 8 values of y.

Warm-start. If the optimization method used to solve the problem (4) can use an initialization of primal or dual variables to reduce the solution time, i.e., take advantage

of a *warm-start*, they can be initialized at an optimal point $x^\star$ for the original problem (1). This can dramatically reduce the time required to solve the N problems. (For more on warm-start methods, see, e.g., Wright (1997), Yildirim and Wright (2002).)

1.5 Connection to stability of solution

When $\epsilon = 0$, the support function evaluation problem (4) is

$$\begin{array}{ll} \text{maximize} & y^T x \\ \text{subject to} & x \in \mathcal{C}, \quad f(x) \leq p^\star. \end{array} \tag{5}$$

Its feasible set is $\mathcal{X}_0$, the set of solutions to the original problem (1).

The support function evaluation problem (4) can be thought of as a perturbed version of (5), where the constraint upper bound is perturbed from its nominal value $p^\star$ to $p^\star + \epsilon$. If $\|y\| = 1$ (which we can assume without loss of generality), the change in optimal value of (4) as ϵ changes is a lower bound on the change in the optimal point $x^\star(y)$, measured in the Euclidean norm. Thus, the size of $\mathcal{X}_\epsilon$ is closely related to the stability of the solution of (4) under perturbation of the constraint bound. For more on general perturbation theory and stability of solutions, see, e.g., (Bonnans and Shapiro 2000; Lucchetti 2006; Rockafellar 1970).

2 Bounding box

The bounding box of $\mathcal{X}_\epsilon$ is the smallest box

$$\mathcal{B} = [l_1, u_1] \times \cdots \times [l_n, u_n]$$

that satisfies $\mathcal{B} \supseteq \mathcal{X}_\epsilon$. The bounding box lower and upper bounds are just the minimum and maximum values that each coordinate x_i can take on, over $\mathcal{X}_\epsilon$. The lower bound l_i is the optimal value of the convex optimization problem

$$\begin{array}{ll} \text{minimize} & x_i \\ \text{subject to} & f(x) \leq p^\star + \epsilon, \quad x \in \mathcal{C}, \end{array}$$

and the upper bound u_i is the optimal value of the convex optimization problem obtained when we change 'minimize' to 'maximize'. Both l_i and u_i can be expressed in terms of the support function of $\mathcal{X}_\epsilon$ as

$$l_i = S_\epsilon(y)(-e_i), \qquad u_i = S_\epsilon(y)(e_i),$$

where $e_i \in \mathbf{R}^n$ is the unit vector whose ith entry is 1. The bounding box corresponds to $\mathcal{P}_{\text{outer}}$ in the case where the vectors $y^{(1)}, \ldots, y^{(N)}$ are chosen to be $\pm e_k$, the unit vectors and their negatives.

Figure 3 shows $\mathcal{X}_\epsilon$ and its bounding box for two problems with $x \in \mathbf{R}^2$, one unconstrained smooth problem, and one with piecewise-linear objective and linear constraints.

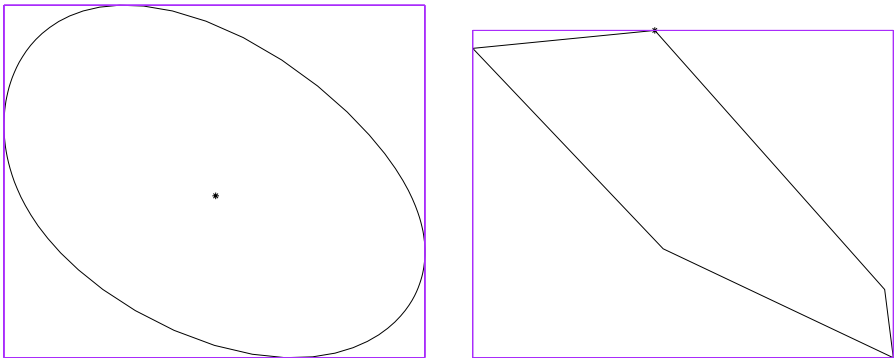

Fig. 3 (Color online) Boundaries of the sets $\mathcal{X}_\epsilon$ (*black*) and the bounding box (*purple*) for an unconstrained problem (*left*) and a nondifferentiable problem with constraints (*right*). The optimal point $x^\star$ is shown as a *star*

3 Approximate quantitative measures

To approximately characterize $\mathcal{X}_\epsilon$, we compute points $x^{(i)} = x^\star(y^{(i)})$, for a set of nonzero $y^{(i)}$, $i = 1, \ldots, N$. These points might include $\pm e_k$, the unit vectors (and their negatives), which allows us to determine (exactly) the bounding box of $\mathcal{X}_\epsilon$. Additional points can be generated randomly, for example from a uniform distribution on the unit sphere in $\mathbf{R}^n$, which can be generated by choosing the components independently from a unit normal distribution, and normalizing y to have length one. (In fact, the normalization is not needed, since scaling $y^{(i)}$ by a positive constant does not affect the corresponding $x^{(i)} = x^\star(y^{(i)})$.)

For general methods of characterizing convex sets using support functions, see, e.g., Fisher et al. (1997), Prince and Willsky (1990), Skiena (1991).

Deviation from $x^\star$. We can estimate the maximum deviation from $x^\star$, in some norm $\|\cdot\|$,

$$\sup\{\|x - x^\star\| \mid x \in \mathcal{X}_\epsilon\}$$

as

$$\max_{i=1,\ldots,N} \|x^{(i)} - x^\star\|.$$

This is a lower bound; it is the same as $\sup\{\|x - x^\star\| \mid x \in \mathcal{P}_{\text{inner}}\}$.

Diameter. The diameter of $\mathcal{X}_\epsilon$,

$$\mathbf{diam}(\mathcal{X}_\epsilon) = \sup_{x,\, \tilde{x} \in \mathcal{X}_\epsilon} \|x - \tilde{x}\|,$$

can be estimated as

$$\mathbf{diam}(\mathcal{X}_\epsilon) \approx \max_{i,j=1,\ldots,N} \|x^{(i)} - x^{(j)}\|.$$

This is a lower bound, and is the same as $\mathbf{diam}(\mathcal{P}_{\text{inner}})$.

Chebyshev center. The Chebyshev center of a set is the center of the largest Euclidean ball that lies in it. (Some authors use Chebyshev center to mean the center of the smallest Euclidean ball that encloses the set). We can estimate the Chebyshev center of $\mathcal{X}_\epsilon$ as the Chebyshev center of $\mathcal{P}_{\text{outer}}$, which can be found by solving the linear program (LP)

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & y^{(i)T}(x_c - x^{(i)}) + r\|y^{(i)}\|_2 \leq 0, \quad i = 1, \ldots, N, \end{array}$$

with variables x_c, the center of the Euclidean ball, and r, its radius. For more on the Chebyshev center, see, e.g., Boyd and Vandenberghe (2004), Sect. 8.5.1.

We can estimate the center of the smallest Euclidean ball that encloses $\mathcal{X}_\epsilon$ as the center of the smallest Euclidean ball that encloses $\mathcal{P}_{\text{inner}}$, which can be found by solving the convex problem

$$\text{minimize} \quad \max_{i=1,\ldots,N} \|x^{(i)} - z\|_2,$$

with variable $z \in \mathbf{R}^n$. This problem is readily reformulated as a convex quadratic program (QP).

Ellipsoidal approximation. We can compute an ellipsoidal approximation of $\mathcal{X}_\epsilon$ in several ways, using the computed boundary points $x^{(1)}, \ldots, x^{(N)}$. We will assume that $\mathcal{X}_\epsilon$ has nonempty interior; if it has empty interior, which means that it lies in an affine space of lower dimension, we can compute an ellipsoidal approximation on that affine space.

The simplest method is to compute their mean and covariance,

$$\bar{x} = (1/N)\sum_{i=1}^{N} x^{(i)}, \qquad \Sigma_{\text{emp}} = (1/N)\sum_{i=1}^{N}(x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^T,$$

and to take our approximation as

$$\mathcal{E}_{\text{emp}} = \{x \mid (x - \bar{x})^T \Sigma^{-1}(x - \bar{x}) \leq \alpha\},$$

where α is chosen to enclose some number (e.g., 95%) of the points.

A more sophisticated method is to compute the *Löwner-John ellipsoid* of $\mathcal{P}_{\text{inner}}$, i.e., the minimum volume ellipsoid $\mathcal{E}_{\text{min}}$ that contains the computed boundary points $x^{(i)}$ (see, e.g., Boyd and Vandenberghe 2004, Sect. 8.4.1; Grötschel et al. 1988; Silverman and Titterington 1980; Sun and Freund 2004). This can be done by solving the convex optimization problem

$$\begin{array}{ll} \text{maximize} & \log\det P \\ \text{subject to} & \|Px^{(i)} + q\|_2 \leq 1, \quad i = 1, \ldots, N, \end{array}$$

with optimization variables $P = P^T \in \mathbf{R}^{n\times n}$ and $q \in \mathbf{R}^n$; the Löwner-John ellipsoid is then

$$\mathcal{E}_{\text{lj}} = \{x \in \mathbf{R}^n \mid \|Pv + q\|_2 \leq 1\}.$$

Another ellipsoidal approximation can be obtained by computing the maximum volume ellipsoid that is contained in the polyhedron $\mathcal{P}_{\text{outer}}$ (see, e.g., Boyd and Vandenberghe 2004, Sect. 8.4.2; Boyd et al. 1994, Sect. 3.) This can be done by solving the convex optimization problem

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \|By^{(i)T}\|_2 + y^{(i)T}\left(d - x^{(i)}\right) \leq 0 \quad i = 1, \ldots, N, \end{array}$$

with optimization variable $B = B^T \in \mathbf{R}^{n\times n}$ and $d \in \mathbf{R}^n$. The maximum volume ellipsoid is then

$$\mathcal{E}_{\text{mve}} = \{Bu + d \mid \|u\|_2 \leq 1\}.$$

An ellipsoidal approximation of $\mathcal{X}_\epsilon$ is useful for several reasons. A large deviation between the center of the approximating ellipsoid and $x^\star$ tells us that $\mathcal{X}_\epsilon$ is not centered at $x^\star$ (as it would be, approximately, for an unconstrained smooth problem, with small ϵ). The semi-axis lengths of the ellipsoid tells us how symmetric $\mathcal{X}_\epsilon$ is. If the semi-axis lengths are not very different, then $\mathcal{X}_\epsilon$ extends roughly the same amount in each direction, from its center; on other hand, if the semi-axis lengths vary considerably, we conclude that $\mathcal{X}_\epsilon$ extends rather different amounts in different directions.

4 Examples

In this section we give four simple numerical examples to illustrate the ideas described above.

4.1 Chebyshev approximation

Our first example is the (unconstrained) Chebyshev approximation problem,

$$\text{minimize} \quad \|Ax - b\|_\infty,$$

with problem data $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$, and variable $x \in \mathbf{R}^n$. Local analysis tells us that if the solution is unique, then this problem has a sharp minimum. This would lead us to guess that $\mathcal{X}_\epsilon$ will be 'small', when ϵ is small.

We consider a problem instance with $m = 100$, $n = 30$, with all entries of A and b chosen randomly from a unit normal distribution. This problem instance has a unique solution $x^\star$, so local analysis tells us that the problem has a sharp minimum. The optimal value is $p^\star = 1.2$. We take $\epsilon = 0.01p^\star$, i.e., we look at the 1%-suboptimal set.

Figure 4 shows $x^\star$ (as stars) and the bounding box of $\mathcal{X}_\epsilon$ (as vertical intervals). We can see that the variation in each x_i over the bounding box is very considerable; for many indices the sign of x_i changes over the bounding box. The fractional variation in x_i, given by $|u_i - l_i|/|x_i^\star|$, ranges between 0.57 and 13.5, with an average value of 2.5. (These can be compared to the difference in objective value obtained, which is 1%.) This shows that the 1%-suboptimal set is relatively large, even though the problem has a sharp minimum. (These two statements are in no way inconsistent.)

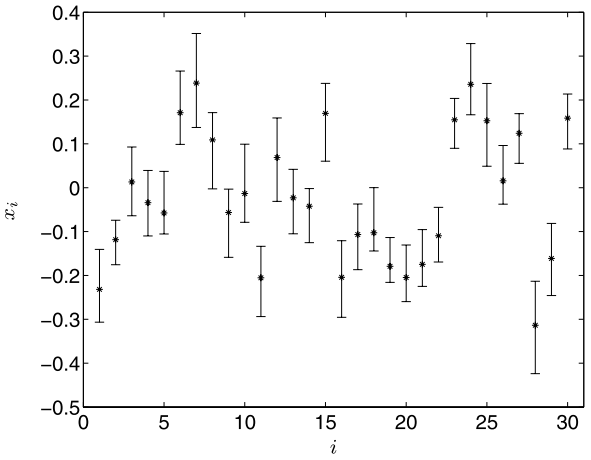

Fig. 4 Solution (shown as *stars*) and bounding box (shown as *vertical intervals*) of 1%-suboptimal set

The figure also shows that $x^\star$ is not too far from the center of the bounding box, suggesting that $x^\star$ is well centered in $\mathcal{X}_\epsilon$.

In addition to the 60 boundary points obtained from the bounding box computation, we compute 940 additional boundary points from randomly chosen directions $y^{(i)}$ (uniform on the unit sphere). From these boundary points we obtain an estimate of (and lower bound on) the maximum deviation (in Euclidean norm) from $x^\star$ of 0.29, around 35% of the (Euclidean) norm of $x^\star$, which is 0.84. This too shows that $\mathcal{X}_\epsilon$ is not small. We estimate the (Euclidean) diameter of $\mathcal{X}_\epsilon$ as 0.45, around half the value of $\|x^\star\|_2$.

We find that the ellipsoids $\mathcal{E}_{\text{emp}}$ and $\mathcal{E}_{\text{lj}}$ are not very different from each other (with proper choice of α in $\mathcal{E}_{\text{emp}}$). These ellipsoids are strongly asymmetric: Two semi-axes are very small (but positive), and the sum of the largest 5 semi-axis lengths is larger than the sum of the remaining semi-axis lengths. The centers of the ellipsoids are not too far from $x^\star$: $\|x^\star - \bar{x}\|_2 = 0.072$, around 9% of $\|x^\star\|_2$, and $\|x^\star - x_{\text{lj}}\|_2 = 0.052$, around 6% of $\|x^\star\|_2$. This indicates that $\mathcal{X}_\epsilon$ is approximately centered around $x^\star$.

Our estimate of (and lower bound on) the diameter of $\mathcal{X}_\epsilon$, as ϵ varies between 0 and $0.02p^\star$, is shown in Fig. 5.

4.2 Minimax estimation

(We are grateful to Yonina Eldar for suggesting this example; see, e.g., Eldar et al. (2007).) We are to estimate a parameter vector $x \in \mathcal{C} \subseteq \mathbf{R}^n$, given a noise-corrupted linear measurement, $b = Ax + v \in \mathbf{R}^m$, where $\mathcal{C}$ is a known closed convex set. We use an unknown-but-bounded model for the noise vector v: It satisfies $\|v\|_2 \leq \eta$, but is otherwise unknown. We will judge the quality of our estimate $\hat{x}$ by the worst-case error, over all values of the true parameter consistent with the measurements,

$$E = \sup\{\|\hat{x} - x\|_2 \mid b = Ax + v,\ \|v\|_2 \leq \eta,\ x \in \mathcal{C}\}.$$

In minimax estimation, we choose $\hat{x}$ to minimize E.

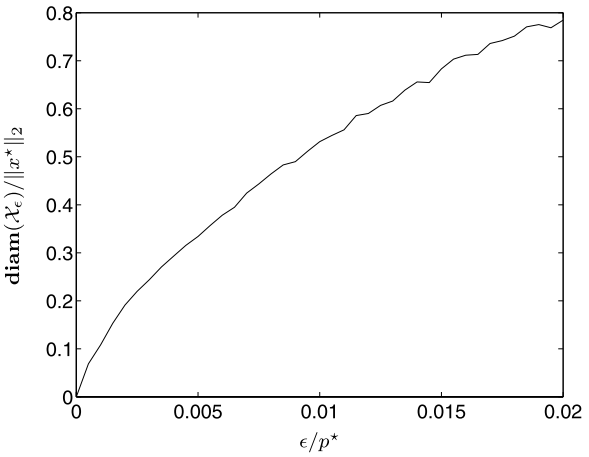

Fig. 5 Estimate of normalized diameter of $\mathcal{X}_\epsilon$ versus $\epsilon/p^\star$

To put this problem in the present framework, we consider the constrained norm minimization problem

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2 \\ \text{subject to} & x \in \mathcal{C}, \end{array} \tag{6}$$

with optimal value $p^\star$. We must have $p^\star \geq \eta$. The associated sublevel set $\mathcal{X}_\epsilon$, for the choice $\epsilon = \eta - p^\star$, is exactly the set of parameter values consistent with the measurement b and the prior information. The bounding box, for example, tells us the exact range of each x_i, over all possible true parameter values.

The minimax estimate is the center of the smallest Euclidean ball that contains $\mathcal{X}_\epsilon$. This can be computed exactly in a few special cases, such as when $\mathcal{C} = \mathbf{R}^n$. In many cases, however, it is difficult to compute the minimax estimate exactly. In these cases we can compute an approximation of the minimax estimate as the center of the smallest Euclidean ball that contains $\mathcal{P}_{\text{inner}}$, which can be found by solving the convex problem

$$\text{minimize} \quad \max_{i=1,\ldots,N} \|x^{(i)} - z\|_2,$$

with variable $z \in \mathbf{R}^n$. (The constraint $z \in \mathcal{C}$ will hold automatically.) This problem, in turn, is readily reduced to a QP. Here $x^{(i)} = x^*(y^{(i)})$, for some set of nonzero directions $y^{(1)}, \ldots, y^{(N)}$.

We illustrate this method with a numerical example, with $n = m = 50$, with the entries of A chosen randomly from a unit normal distribution. We take $\mathcal{C} = \{x \mid \|x\|_\infty \leq 1\}$, and $\eta = 5$. The true parameter value x is chosen from a uniform distribution on $\mathcal{C}$. Finally, v is chosen from a uniform distribution on the η-sphere. (Recall, however, that the minimax problem itself does not involve any probability distributions.)

For this problem instance we compute the minimum norm estimate $\hat{x}_{\text{mn}}$, which is the solution of (6). We also compute an approximation of the minimax estimate $\hat{x}_{\text{amm}}$ using the scheme outlined above, using $y^{(i)}$ as $\pm e_j$, and an additional 900 randomly chosen $y^{(i)}$.

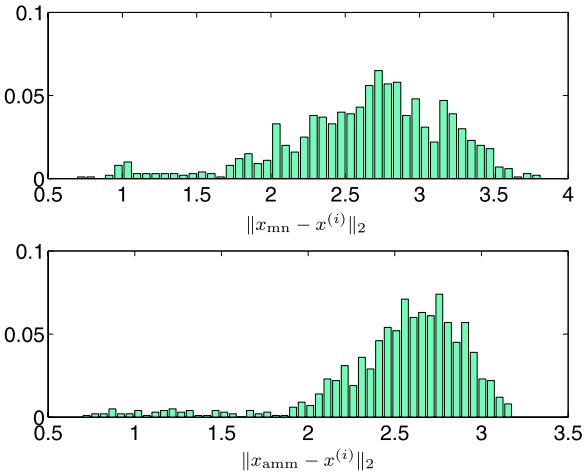

Fig. 6 Distributions of $\|x_{\mathrm{mn}} - x^{(i)}\|_2$ (*top*) and $\|x_{\mathrm{amm}} - x^{(i)}\|_2$ (*bottom*) over $i = 1, \ldots, 1000$

We can then compare (our estimate of) the worst-case error E for $\hat{x}_{\mathrm{mn}}$ and $\hat{x}_{\mathrm{amm}}$, using

$$\hat{E} = \max_{i=1,\ldots,1000} \|\hat{x} - x^{(i)}\|_2.$$

The results are $\hat{E} = 3.82$ for $\hat{x}_{\mathrm{mn}}$, and $\hat{E} = 3.18$, for $\hat{x}_{\mathrm{amm}}$. The true errors are

$$\|x_{\mathrm{mn}} - x\|_2 = 1.78, \qquad \|x_{\mathrm{amm}} - x\|_2 = 1.40,$$

where x is the true parameter. Thus, our (approximate) minimax estimate gives a substantial decrease in (approximate) worst-case error, as well as in true error.

Figure 6 shows the distributions of $\|\hat{x} - x^{(i)}\|_2$, over $i = 1, \ldots, 1000$, for the two estimates. (Again we remind the reader that the minimax estimation problem itself does not involve any probability distributions; this is only to illustrate the range of values of $\|\hat{x} - x^{(i)}\|_2$, over our sample points $x^{(i)}$.)

4.3 Optimal control

Our next example is a discrete-time finite horizon optimal control problem. The system dynamics are given by

$$x(t+1) = Ax(t) + Bu(t), \quad t = 0, 1, \ldots, t_f - 1,$$

where t represents time (or period), t_f is the horizon (or final time), $x(t) \in \mathbf{R}^n$ is the state of the system, and $u(t) \in \mathbf{R}^m$ is the control input. The initial state x_0 is given, and we require the final state $x(t_f)$ to be zero (this is called the *regulation problem*). This is equivalent to

$$\sum_{t=0}^{t_f-1} A^{t_f-t-1} Bu(t) = -A^{t_f} x_0. \tag{7}$$

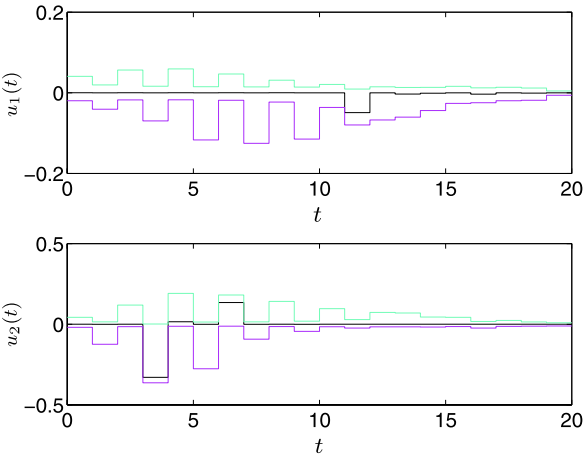

Fig. 7 Optimal input, and lower and upper bounds over 5%-suboptimal trajectories

Using the ℓ_1-norm of the input (which is a typical simple model for total fuel usage) as objective, our problem is

$$\begin{array}{ll} \text{minimize} & \|u\|_1 = \sum_{t=0}^{t_f-1} \|u(t)\|_1 \\ \text{subject to} & \sum_{t=0}^{t_f-1} A^{t_f-t-1} B u(t) = -A^{t_f} x_0, \end{array}$$

with variables $u = (u(0), \ldots, u(t_f - 1))$, and problem data A, B, x_0, and t_f. Local analysis tells us that when the optimal point is unique, this problem has a sharp minimum.

We now consider a specific problem instance with $n = 8$, $m = 2$, $t_f = 20$, and all coefficients in the problem data drawn from a unit normal distribution. We then scale A so that its spectral radius is 1.05, i.e., the system is slightly unstable. For this problem instance, the optimal point $u^\star$ is unique, and the optimal cost is $p^\star = 0.54$. Thus, for this example too, the problem has a sharp minimum. We take $\epsilon = 0.05 p^\star$, i.e., we look at the set of 5%-suboptimal input trajectories. To avoid confusion, we will denote the ϵ-suboptimal set as $\mathcal{U}_\epsilon$, since the variable here is u.

Figure 7 shows the optimal input trajectory $u^\star$, and the bounding box for $\mathcal{U}_\epsilon$. We see immediately that $\mathcal{U}_\epsilon$ is large; for most values of t, $u_i(t)$ can vary over a relatively large interval (and certainly far larger than 5%). Indeed, the upper bounds are all positive, while the lower bounds are negative. This means that we can specify the sign of any particular variable $u_i(t)$, and still find a 5% suboptimal input trajectory.

Using the 80 boundary points obtained from the bounding box computation and 920 additional ones computed from randomly chosen directions $y^{(i)}$, we estimate (lower bound) the maximum deviation from $u^\star$ over $\mathcal{U}_\epsilon$ as 1.06 in ℓ_1-norm, almost twice the value of $\|u^\star\|_1$. This is quite surprising, since the triangle inequality tells us that any point in $u \in \mathcal{U}_\epsilon$ must satisfy

$$\|u - u^\star\|_1 \leq \|u\|_1 + \|u^\star\|_1 \leq p^\star(2 + \epsilon),$$

which means that $\|u - u^\star\|_1$ cannot be more than $2.05\|u^\star\|_1$. Again we see that there are 5%-suboptimal input trajectories that are very different from the optimal input trajectory.

Using the same 1000 points, we estimate the diameter of $\mathcal{U}_\epsilon$ as 1.13 in ℓ_1-norm.

We also compute the ellipsoid $\mathcal{E}_{\text{emp}}$ (on the 32-dimensional affine set defined by (7)). On this lower dimensional space the ellipsoid extends different amounts in different directions. The sum of its largest 3 semi-axis lengths is larger than the sum of the remaining semi-axis lengths. The distance between $u^\star$ and $\bar{u}$, the center of $\mathcal{E}_{\text{emp}}$, is 0.1975 in ℓ_2-norm (0.5074 in ℓ_1-norm).

Our conclusion that $\mathcal{U}_\epsilon$ is large can be used in several ways. It suggests that, by accepting a small increase in fuel usage, we might see a substantial improvement in some secondary objective. For example, we can pick, from $\mathcal{U}_\epsilon$, an input trajectory that is maximally smooth.

4.4 Portfolio optimization

We consider a simple portfolio allocation problem with n assets held over a period of time. We let x_i denote the fraction that we invest in asset i. We do not allow short positions, so the portfolio allocation x must satisfy $x \geq 0$, $\mathbf{1}^T x = 1$.

The vector of asset (percentage) returns is the random variable $r \in \mathbf{R}^n$, with known mean $\mu = \mathbf{E}r$ and covariance $\Sigma = \mathbf{E}(r - \mu)(r - \mu)^T$. The overall return on the portfolio is the scalar random variable $r^T x$, which has mean $\mu^T x$ and standard deviation $(x^T \Sigma x)^{1/2}$.

The goal is to find a portfolio allocation x that minimizes the portfolio return standard deviation (i.e., risk), subject to a minimum expected return $r_{\min}$, and the allocation constraints. This is the convex optimization problem

$$\begin{array}{ll} \text{minimize} & (x^T \Sigma x)^{1/2} \\ \text{subject to} & \mu^T x \geq r_{\min}, \\ & x \geq 0, \quad \mathbf{1}^T x = 1, \end{array}$$

with variable x and problem data μ, Σ, and $r_{\min}$. For more on the portfolio optimization problem, see, e.g., Markovitz (1952); Boyd and Vandenberghe (2004), Sect. 4.4.

We consider the case where r is described by a single-factor model, i.e., $r_i = \mu_i + \beta_i w + v_i$, for $i = 1, \ldots, n$. Here β_i's are the factor loadings, w is the (market) factor, and v_i's represent the residual risks. We take w and v_i's to be independent Gaussian zero-mean random variables, with respective standard deviations σ_{m} and σ_i. We can think of σ_{m} as the market-related standard deviation and σ_i as the firm-specific standard deviation. Under this model, the return covariance matrix is

$$\Sigma = \sigma_{\text{m}}^2 \beta\beta^T + \mathbf{diag}(\sigma_1^2, \ldots, \sigma_n^2),$$

where $\beta = (\beta_1, \ldots, \beta_n)$.

We now consider a particular problem instance with $n = 10$ assets. We take market standard deviation $\sigma_{\text{m}} = 20$; β_i are chosen uniformly on $[0.3, 1]$ for $i = 2, \ldots, n$, and

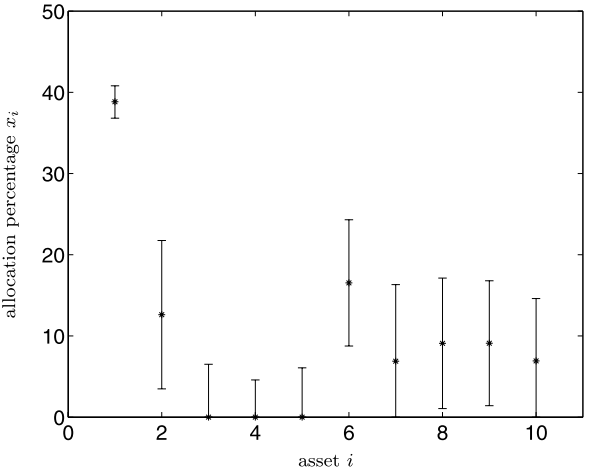

Fig. 8 Optimal portfolio allocation (shown as *stars*) and bounding box (shown as *vertical intervals*) for the $0.02p^\star$-suboptimal set

σ_i are chosen uniformly on $[0, 20]$ for $i = 2, \ldots, n$. We take the first asset to be risk-free (i.e., $\beta_1 = 0$, $\sigma_1 = 0$). The mean returns of the assets are given by

$$\begin{aligned}\mu_i &= \mu_{\mathrm{rf}} + 0.4(\mathbf{E}(\beta_i w + v_i)^2)^{1/2} \\ &= \mu_{\mathrm{rf}} + 0.4\left(\beta_i^2\sigma_{\mathrm{m}}^2 + \sigma_i^2\right)^{1/2}, \quad i = 1, \ldots, n,\end{aligned}$$

where $\mu_{\mathrm{rf}} = 5$ is the risk-free return. (Here, σ_{m}, σ_i, and μ_i are all expressed in percentage points, and the constant 0.4 is the reward-to-risk ratio.) The 10 assets have risk (standard deviation) ranging from 0 (for the risk-free asset n) to 22.9, and mean returns ranging from 5 (the risk-free return) to 14.15. We re-order the assets by increasing risk (and return). We take the required minimum expected return to be $r_{\min} = 10$.

The optimal portfolio, which uses only 6 of the 10 assets, achieves the minimal risk (standard deviation) $p^\star = 7.84$. We will examine $\mathcal{X}_\epsilon$ for $\epsilon = 0.02p^\star = 0.16$, i.e., we look at the set of portfolios that result in standard deviation that is no more than 8, i.e., 1.02 times the optimal standard deviation. Figure 8 shows the optimal asset allocation, and the bounding box of $\mathcal{X}_\epsilon$. We see that the set of suboptimal portfolio allocations is large; there seems to be considerable variation among the suboptimal portfolio allocations. One exception is the allocation in the risk-free asset, which only varies between 36.8% and 40.8% over all suboptimal asset allocations. But allocations of some other assets exhibit large deviations around the optimal allocation: for example, the allocation of asset 7 varies between 0 and 16.3% over the suboptimal portfolios.

Using the boundary points from the bounding box computations and 1000 additional ones computed from randomly chosen directions $y^{(i)}$, we estimate the maximum deviation (in Euclidean norm) from $x^\star$ as 0.1028, around 22% of the (Euclidean) norm of $x^\star$, which is 0.47. We also estimate the (Euclidean) diameter of the suboptimal set to be 0.203, around 43% of $\|x^\star\|_2$.

We also compute the ellipsoids $\mathcal{E}_{\mathrm{emp}}$ and $\mathcal{E}_{\mathrm{lj}}$ (in the 9-dimensional affine set defined by $\mathbf{1}^T x = 1$). On this lower dimensional space, the ellipsoids extend roughly

the same amount in 8 directions and are quite flat in the remaining direction. Furthermore, the (Euclidean) distance between $x^\star$ and $\bar{x}$ is 0.0056, i.e., 1.2% of $\|x^\star\|_2$, and the (Euclidean) distance between $x^\star$ and the center of $\mathcal{E}_{\rm lj}$ is 0.0139, i.e., around 3% of $\|x^\star\|_2$. This strongly indicates that $\mathcal{X}_\epsilon$ is well centered around $x^\star$.

Our observation that the set of suboptimal asset allocations is large (at least in one direction) suggests that we may be able to substantially improve a secondary objective, such as the trading costs incurred, given a starting allocation, while increasing the risk only slightly.

4.5 Constrained maximum likelihood estimation

Our last example is a constrained maximum likelihood estimation problem. We wish to estimate a vector $x \in \mathbf{R}^n$, which is known to satisfy m linear inequalities $Fx \preceq g$, where $F \in \mathbf{R}^{m\times n}$ and $g \in \mathbf{R}^m$. We are given a noise-corrupted linear measurement of x, given by $y = Ax + v$, where $A^{p\times n}$ and $v \in \mathbf{R}^p$ is a zero-mean Gaussian random vector, with (positive definite) covariance Σ.

The likelihood function is

$$L(x) = (2\pi)^{n/2}|\det(\Sigma)|^{-1/2}\exp\bigl(-(1/2)(y-Ax)^T\Sigma^{-1}(y-Ax)\bigr).$$

The log-likelihood function is

$$l(x) = \log(L(x)) = -(1/2)\|\Sigma^{-1/2}(y-Ax)\|_2^2 + c,$$

where $c = -(n/2)\log(2\pi) - (1/2)\log|\det(\Sigma)|$ does not depend on x. The maximum likelihood estimate $x_{\rm ml}$ is found by maximizing the likelihood function (or, equivalently, the log-likelihood function) subject to the linear inequality constraints. Thus, $x_{\rm ml}$ is the solution of the QP

$$\begin{array}{ll} \text{maximize} & -(1/2)\|\Sigma^{-1/2}(y-Ax)\|_2^2 \\ \text{subject to} & Fx \preceq g, \end{array}$$

with variable $x \in \mathbf{R}^n$, and problem data $(\Sigma^{-1/2}y)$, $(\Sigma^{-1/2}A)$, F, and g. The ϵ-suboptimal set $\mathcal{X}_\epsilon$ is the set of parameter values with log-likelihood within ϵ of the maximum log-likelihood value. These correspond to parameter values that are almost as plausible as the maximum likelihood parameter $x_{\rm ml}$, given the observation.

We can give a more specific interpretation of $\mathcal{X}_\epsilon$ in terms of a hypothesis test. Let us fix a particular $z \in \mathbf{R}^n$. Given the observation y, which is generated either as $y = Ax_{\rm ml} + v$, or as $y = Az + v$, we are to guess which of the two distributions generated y. Thus we have two hypotheses: y came from the maximum likelihood distribution, or from the distribution defined by z. We have two associated error probabilities: $p_{\rm fp}$, which is the probability that we guess $x_{\rm ml}$, under the distribution associated with z; $p_{\rm fn}$, which is the probability that we guess z, under the distribution associated with $x_{\rm ml}$. The estimator that minimizes the weighted cost

$$w_{\rm fp}p_{\rm fp} + w_{\rm fn}p_{\rm fn},$$

where w_{fp} and w_{fn} are positive weights, is a likelihood ratio threshold test; it has the form

$$\text{guess } z \text{ when } \quad z \in \mathcal{X}_\epsilon,$$

where $\epsilon = \max\{0, \log(w_{\mathrm{fp}}/w_{\mathrm{fn}})\}$. (If $w_{\mathrm{fp}} \leq w_{\mathrm{fn}}$, we always guess x_{ml}.) This estimator equalizes the weighted cost terms, i.e., we have

$$p_{\mathrm{fp}}/p_{\mathrm{fn}} = w_{\mathrm{fn}}/w_{\mathrm{fp}} = e^{-\epsilon}.$$

For example with $\epsilon = 0.1$, $\mathcal{X}_\epsilon$ is the set of parameter values that would be chosen over x_{ml} in a hypothesis test in which we weight false positives (choosing x_{ml}) a fraction $e^{0.1}$ times more than false negatives (choosing z).

We now turn to a more specific case: estimating a nondecreasing signal given a noise-corrupted moving average. Our goal is to estimate a scalar signal $x(t)$, for $t = 1, 2, \ldots, n$, which is known to be nondecreasing, i.e., $x(1) \leq x(2) \leq \cdots \leq x(n)$. We are given a noise-corrupted moving sum of x, given by

$$y(t) = \sum_{\tau=0}^{\max\{k-1,t-1\}} x(t-\tau) + v(t), \quad t = 1, \ldots, n,$$

where $v(t)$ are independent $\mathcal{N}(0, 1)$ random variables, and k is the width of the moving-sum window. We can write this as $y = Ax + v$, where x, y and $v \in \mathbf{R}^n$ and $A \in \mathbf{R}^{n\times n}$ is the lower triangular Toeplitz matrix given by

$$A_{ij} = \begin{cases} 1 & 0 \leq j - i \leq k-1 \\ 0 & \text{otherwise.} \end{cases}$$

The monotonicity constraints on x can be expressed as a set of linear inequalities, written in matrix form as $Fx \leq 0$, where F is the backward difference matrix

$$D = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix} \in \mathbf{R}^{(n-1)\times n}.$$

(Zero elements are not shown.)

We now consider a particular problem instance, with $n = 25$, $k = 4$, and y as one sample from the distribution $A\bar{x} + v$, where the (true) signal $\bar{x}$ is a randomly generated nondecreasing vector in $\mathbf{R}^n$. The maximum log-likelihood is found to be $p^\star = -32.9$. We take $\epsilon = 0.1$.

Figure 9 shows the bounding box around $\mathcal{X}_\epsilon$, as well as the maximum likelihood signal x_{ml}. We see that for each x_i, $u_i - l_i$ is around 3, except at the right end, where fewer measurements of the variables are available. Any value in the interval $[l_1, u_i]$ can be defended as a plausible value for x_i, since an x with this value would be chosen in a hypothesis test against x_{ml}, with relative weight $e^{0.1} = 1.052$ of false positives.

Using the 60 boundary points from the bounding box computations and 940 additional ones computed from randomly chosen directions $y^{(i)}$, we find that the maximum deviation (in Euclidean norm) from $x^\star$ is 4.4450, around 8% of the (Euclidean)

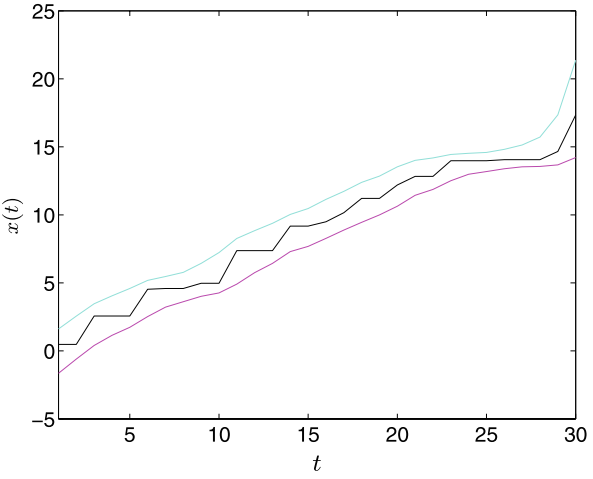

Fig. 9 (Color online) Maximum likelihood estimate (shown in *black*) and bounding box, shown as lower (in *purple*) and upper (in *cyan*) bounds

norm of $x^\star$, which is 55.5051. We also estimate the (Euclidean) diameter of the suboptimal set to be 7.1780, around 13% of $\|x^\star\|_2$. This indicates that $\mathcal{X}_\epsilon$ is relatively small.

We also compute the ellipsoids $\mathcal{E}_{\text{emp}}$ and $\mathcal{E}_{\text{lj}}$ and discover that they extend different amounts in different directions. For both ellipsoids, the sum of the largest 8 semi-axis lengths is larger than the sum of the remaining semi-axis lengths. Furthermore, the (Euclidean) distance between $x^\star$ and $\bar{x}$ is 1.2684, i.e., around 2% of $\|x^\star\|_2$, and the (Euclidean) distance between $x^\star$ and the center of $\mathcal{E}_{\text{lj}}$ is 1.0116, i.e., around 1.8% of $\|x^\star\|_2$. This strongly suggests that $\mathcal{X}_\epsilon$ is well centered around $x^\star$.

5 Conclusions

For an unconstrained problem with smooth objective, the suboptimal set is easily characterized: For small enough ϵ, it is approximately ellipsoidal, centered at the optimal point, with shape determined by the Hessian evaluated at the optimal point. In other cases, however, the suboptimal set can be, and often is, very different. The examples above show that the suboptimal set can be much larger than one might expect: Allowing just a few percent suboptimality can allow individual components of the variable to vary considerably, or even change sign. This is observed even in problems with a sharp minimum, for which local analysis suggests that the suboptimal set is small. Of course the opposite can also occur: The suboptimal set can be small, even when local analysis suggests that it should be large. Our examples show that carrying out at least some analysis or exploration of the suboptimal set, for example, computing its bounding box, can be quite informative in a practical setting.

The general idea of characterizing a convex set via its support function, as well as the various approximations we have described, are well known. It is an elementary exercise to compute the bounding box of the ϵ-suboptimal set. But we have not seen these ideas described, or suggested, in the context of suboptimal set exploration.

Acknowledgements We thank Andrea Montanari and Thomas Cover for their helpful discussions of the maximum likelihood example, and Yonina Eldar for suggesting the minimax estimation example, as well as general suggestions.

References

Ben-Tal A, Nemirovski A (2001) Lectures on modern convex optimization: Analysis, algorithms, and engineering applications. MPS-SIAM Series on Optimization

Bertsekas D, Nedić A, Ozdaglar A (2003) Convex analysis and optimization. Athena Scientific, Belmont

Bonnans J, Shapiro A (2000) Perturbation analysis of optimization problems. Springer, Berlin

Borwein J, Lewis A (2000) Convex analysis and nonlinear optimization: theory and examples. Springer, Berlin

Boyd S, El Ghaoui L, Feron E, Balakrishnan V (1994) Linear matrix inequalities in systems and control theory. SIAM, Philadelphia

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge

Eldar Y, Beck A, Teboulle M (2007) A minimax Chebyshev estimator for bounded error estimation. CCIT Report, 617, EE Department, Technion

Fisher N, Hall P, Turlach B, Watson G (1997) On the estimation of a convex set from noisy data on its support function. Am Stat Assoc 92:84–91

Grötschel M, Lovasz L, Schrijver A (1988) Geometric algorithms and combinatorial optimization. Springer, Berlin

Lucchetti R (2006) Convexity and well-posed problems. Springer, Berlin

Markovitz H (1952) Portfolio selection. J Finance 7(1):77–91

Nesterov Y (2003) Introductory lectures on convex optimization: a basic course. Springer, Berlin

Polyak B (1987) Introduction to optimization. Optimization Software, New York

Prince J, Willsky A (1990) Reconstructing convex sets from support line measurements. IEEE Trans Pattern Anal Mach Intell 12(4):377–389

Rockafellar R (1970) Convex analysis. Princeton University Press, Princeton

Silverman BW, Titterington DM (1980) Minimum covering ellipses. SIAM J Sci Stat Comput 1(4):401–409

Skiena S (1991) Probing convex polygons with half-planes. J Algorithms 12(3):359–374

Sun P, Freund R (2004) Computation of minimum-volume covering ellipsoids. Oper Res 52(5):690–706

Wright S (1997) Primal-dual interior-point methods. SIAM, Philadelphia

Yildirim E, Wright S (2002) Warm-start strategies in interior-point methods for linear programming. SIAM J Optim 12(3):782–810