Convex Optimization

Problem set 1

Due Monday, March 8th

Convexity

1. Consider the closed convex set $B_1 = \{x \in \mathbb{R}^n | \|x\|_1 = \sum_i |x_i| \leq 1\}$. This is the unit ball of the ℓ_1 norm.

   (a) Show that B_1 is a polyhedron by explicitly expressing it as an intersection of halfspaces. How many halfspaces ("facets") are required in order to express B_1?

   (b) Explicitly express B_1 as a convex hull of a finite number of points. How many points ("vertices") are required in this characterization?

   (c) Contrast this with the ℓ_∞ unit ball, $B_\infty = \{x \in \mathbb{R}^n | \|x\|_\infty \leq 1\}$. How many halfspaces are required in order to express B_∞ as an intersection of halfspaces? How many points are required in order to express B_∞ as a convex hull?

   (d) For each point $\hat{x}$ on the boundary of B_1, identify the set of all supporting hyperplanes of B_1 at $\hat{x}$ explicitly. For each such $\hat{x}$, what is the dimensionality of this set?

2. Consider a polyhedron $C = \text{conv}\{v_1, \ldots, v_k\} \subset \mathbb{R}^n$ and a convex function $f : \mathbb{R}^n \to \mathbb{R}$.

   (a) Prove that a maximum of f over C is achieved at one of the vertices v_i. (Hint: assume the statement is false and use Jensen's inequality). Is it possible that the maximum is also achieved at an interior point?

   (A generalization of the above is that a maximum of a function over a closed and bounded convex set is achieved at an extreme point, i.e. a point which is not a convex combination of other points in the set).

   (b) Use the above to conclude that the *minimum* of a linear objective over the polyhedron C is always achieved at one of the vertices v_i.

3. In this problem we will define strong convexity more generally then it is defined by Boyd and Vandenberghe (Section 9.1.2). In particular, we will consider a definition that is valid also for non-differentiable functions.

   Definition: A function $f : \mathbb{R}^n \to \mathbb{R}$ is m-strongly convex if for every $x, y \in \mathbb{R}^n$ and every $\theta \in [0, 1]$:

$$f((1-\theta)x + \theta y) \leq (1-\theta)f(x) + \theta f(y) - \frac{m}{2}\theta(1-\theta)\left\|x - y\right\|_2^2$$

   (a) Prove that a continuously differentiable function f is m-strongly convex if and only if for every $x, y \in \mathbb{R}^n$,

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\left\|y - x\right\|_2^2.$$

   This generalizes the first order characterization of convexity (Section 3.1.3).

   (b) Prove that a twice continuously differentiable function f is m-strongly convex if and only if its domain and convex and for every $x \in \mathbb{R}^n$, all eigenvalues of the Hessian at x are greater or equal to m, i.e.:

$$\nabla^2 f(x) \succcurlyeq mI.$$

   This generalizes the second order characterization of convexity (Section 3.1.4) and is the definition used in Section 9.1.2.

   (c) Provide an example of a function that is strongly convex but not everywhere differentiable.

   (d) Let f be a m-strongly convex function, and x^* an optimum for $\min_{x \in \mathbb{R}^n} f(x)$. Prove that for any point $x \in \mathbb{R}^n$:

$$f(x) \geq f(x^*) + \frac{m}{2}\left\|x - x^*\right\|_2^2.$$

   Conclude that the optimum is unique and that any ϵ-suboptimal point must be close to the optimum. Provide an explicit upper bound on $\left\|x - x^*\right\|_2$ for an ϵ-suboptimal x. (Note that if f is convex but not strongly convex, ϵ-suboptimal points can be arbitrarily far away from the closest optimum).

Recommended review exercises from Boyd and Vandenberghe (please do not turn these in—they will *not* be graded): 2.12, 2.15, 3.6, 3.16, 3.18, 3.24, 3.26.

Unconstrained Optimization

4. In this problem we will consider gradient descent with predetermined step sizes. That is, instead of determining $t^{(k)}$ by a linesearch method using the objective function, the current iterate $x^{(k)}$ and the descent direction $\triangle x^{(k)}$, it will be set to some pre-determined sequence.

   (a) Show that gradient descent with a fixed step size of $t^{(k)} = 1$ can lead to a sequence of iterates that does not converge to the optimum value. In particular, show an explicit strongly-convex function $f(\cdot)$ and initial point $x^{(0)}$, such that we can have $\inf_k x^{(k)} > \inf_x f(x)$.

   (b) Next, show that even setting the step size according to the function does not help: Show a strongly convex function $f(\cdot)$ such that for any fixed step size t, there exists an initial point $x^{(0)}$, for which the sequence of iterates generated by gradient descent with $t^{(k)} = t$ satisfies $\inf_k x^{(k)} > \inf_x f(x)$.

   (c) Consider a predetermined sequence of iterates such that $\lim_{k\to\infty} t^{(k)} = 0$ and $\sum_{k=0}^{\infty} t^{(k)} = \infty$ and a twice continuously differentiable convex function $f : \mathbb{R}^n \to \mathbb{R}$ with a bounded Hessian $\nabla^2 f(x) \preccurlyeq MI$. **You may also assume that the sublevel sets of the function are bounded**. Prove that gradient descent on $f(\cdot)$ with steps of sizes $t^{(k)}$ converges to the optimal value: $\lim_{k\to\infty} f(x^{(k)}) = \inf_x f(x)$. (**Hint:** First establish that for large enough k, $f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{1}{2} t^{(k)} \left\| \nabla f(x^{(k)}) \right\|^2$. Next bound the norm of the gradient in terms of the suboptimality. Now establish that for any ϵ, we will reach an ϵ-suboptimal point after a finite number of iterations.)

   Note that we do not require that f is strongly convex, we only bound the Hessian from above. If the step sizes decreases slowly enough, namely as $t^{(k)} = 1/\sqrt{k}$, then it is not even necessary to bound the Hessian from above, nor assume $f(\cdot)$ is twice continuously differentiable. However, convergence achieved with such predetermined step sizes could be much slower, with the number of iterations increasing polynomially, rather then logarithmically, with $1/\epsilon$.

5. Boyd and Vandenberghe Problems 9.17(c) and 9.18.

6. We defined self-concordance of a scalar function using the condition $|f'''(t)| \leq 2(f''(t))^{3/2}$. The constant 2 in this definition is arbitrary, and this definition depends on the scaling of the function $f(t)$. In this problem, we will consider κ-self concordant functions: For $\kappa > 0$, we say that a convex scalar function is κ-self concordant iff for all t:

$$|f'''(t)| \leq \kappa(f''(t))^{3/2}$$

   We say $f : \mathbb{R}^n \to \mathbb{R}$ is κ-self concordant if $t \mapsto f(x + tv)$ is κ-self-concordant for all $x, v \in \mathbb{R}^n$.

   (a) Show that $f(t) = -\log t$ is self-concordant under the simpler definition, but $f(t) = -\frac{1}{2}\log t$ is not. For what value of κ is it κ-self concordant?

   (b) What functions are κ-self concordant for all $\kappa > 0$?

(c) If a function $f(\cdot)$ is κ-self concordant, for what scaling constant $c > 0$ is the functions $\tilde{f}(x) = cf(x)$ self-concordant under the simpler definition?

(d) Consider unconstrained minimization of a strictly convex function $f(\cdot)$ that is κ-self concordant. Provide a bound on the suboptimality $f(x) - p^*$ in terms of κ and the Newton decrement $\lambda(x)$, for small enough values of $\lambda(x)$ (be sure to specify for what values of $\lambda(x)$ the bound is valid). (See Section 9.6.3 of Boyd and Vandenberghe. Hint: how do the suboptimality and the Newton decrement scale when the function is scaled?).

7. **Steepest Descent** In this problem, we will define the direction of Steepest Descent with respect to a norm, and investigate it for several specific norms. For a norm $\|\cdot\|$ on $\mathbb{R}^n$, a direction of steepest descent of $f : \mathbb{R}^n \to \mathbb{R}$ at x with respect to $\|\cdot\|$ is given by:

$$\begin{aligned}\triangle x &\stackrel{\text{def}}{=} \arg\min_{\|v\|=1} \nabla f(x)^T v = \arg\min_{\|v\|=1} \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} \\ &\approx \lim_{t\to 0} \arg\min_{\|v\|=1} \frac{f(x+tv) - f(x)}{t} = \lim_{t\to 0} \arg\min_{\|v\|=t} f(x+v)\end{aligned}$$

where the $\approx$ should be interpreted only as an intuitive correspondence, and so the second line only as a rough intuition (e.g. the $\arg\min$ might have multiple minimizers, making the limit not properly defined. It is even possible that some directions of steepest descent, as defined by the first line, are not the limit point for any sequence of minimizers in the second line). But roughly speaking, $\triangle x$ is the direction v in which the greatest decrease in $f(x)$ can be achieved subject to an infinitesimally small constraint on $\|v\|$.

A steepest descent method uses the direction of steepest descent as a descent direction at each iteration, taking a step in this direction with a stepsize chosen by some linesearch method.

(a) Show that the direction of steepest descent for the Euclidean ℓ_2 norm is $\triangle x = -\nabla f(x)$, and thus steepest descent with respect to the Euclidean norm is just gradient descent.

(b) For $H \succ 0$, consider the norm $\|x\| = x^T H x$. What is the direction of steepest descent with respect to this norm?

(c) What are the directions of steepest descent with respect to the ℓ_1 norm $\|x\| = \sum_i |x_i|$? Explain how to find these directions, and think of how a steepest descent method w.r.t. the ℓ_1 norm would proceed.

(d) What are the directions of steepest descent with respect to the ℓ_∞ norm $\|x\| = \max_i |x_i|$? Explain how to find these directions, and think of how a steepest descent method w.r.t. the ℓ_∞ norm would proceed.