California Institute of Technology **HSS Division**
Crash-Course on Convex Optimization **Fall 2011-12**
Ec 133
Guilherme Freitas

In this text, we will study the following ***basic problem***:

$$\begin{aligned} \underset{x \in C}{\texttt{maximize}} \quad & f(x) \\ \texttt{subject to} \quad & h_l(x) = 0 \quad l = 1, \ldots, m \\ & g_j(x) \leqslant 0 \quad j = 1, \ldots, k \end{aligned}$$

where $C \subset \mathbb{R}^n$ is a convex set, $f : C \to \mathbb{R}$ is a continuously differentiable concave function, for all $l \in \{1, \ldots, m\}$, $h_l : \mathbb{R}^n \to \mathbb{R}$ is an affine function,[1] and for all $j \in \{1, \ldots, k\}, g_j : \mathbb{R}^n \to \mathbb{R}$, is a continuously differentiable convex function. In other words, this is a smooth ***convex optimization problem*** with m 'linear' (or more properly, affine) constraints and k nonlinear constraints.[2] A point $x \in \mathbb{R}^n$ is ***feasible*** when $x \in C$ and the constraints $h(x) = 0$ and $g(x) \leqslant 0$ are satisfied. A feasible point x is a ***solution*** when for any other feasible point y we have $f(x) \geqslant f(y)$. In that case, we also say that x is ***optimal***.[3]

We make two main points in these notes:

1. As long as the constraints of the problem satisfy certain regularity conditions often encountered in practice, *the gradient of the objective function evaluated at a solution will be a linear combination of the gradients of the active constraints; the coefficients of such a linear combination are called Lagrange multipliers*. Actually, we can do better than that: we can also determine that the multipliers of the inequality constraints are nonnegative (see the sections on the visual and penalty intuition).

2. It is possible to cast some nonconvex problems as convex optimization problems. This allow us to leverage all the machinery from convex optimization in these situations. We will illustrate the concept with two specific techniques that are very easy and often applicable: *convex relaxation*, and the *epigraph transformation*.

---

[1] Affine means a function of the form $h_l(x) = \langle a_l, x \rangle + b_l$, where $a_l, b_l \in \mathbb{R}^n$.

[2] Most texts will talk about minimizing a convex f, which is the same as maximizing a concave $-f$. In this course, the natural formulation of most problems is to maximize some function, hence the formulation above.

[3] It is common to see the terms "feasible solution" instead of feasible point and "optimal solution" instead of solution.

In the following, I go over what you need to know in order to understand this material properly, some intuition and motivation. Only after that I get to the main results. If you want to cut to the chase, go the section *Main Results* in the end, but I strongly recommend reading the sections on the intuition behind the problem. I tried to increase the usefulness of these notes by pointing to close/useful directions in both the theory of convex/constrained optimization and practical solution of convex problems. I hope they will not get in the way of those who want just the minimum material for Ec 133, that is the corollary in the main results section.

Read This Before You Proceed

I assume you have taken a multivariate calculus course and that your basic linear algebra is not too rusty. Additionally, make sure you understand the following:

- Definition of convex set.
- Definition of a convex function.
- Definition of a concave function.
- Definition of upper and lower contour set.
- The lower contour sets of a convex function are convex sets; the upper contour sets of a concave function are convex sets. Note that it is *not* the case that that all functions that have convex upper contour sets are concave functions. Think of the bell curve.[4]
- Definition of a constrained optimization problem, and related terminology: parameters, variables, constraints, active (or binding, which is the same) constraints, inactive (or not binding) constraints, objective function, local minimum, global minimum, local minimizer, global minimizer, first-order conditions (think "derivative equals zero"), second-order conditions (think "second derivative has to be negative for a maximum").
- A vector $v \in \mathbb{R}^n$ divides the space in two half-spaces, the positive one and the negative one. For every vector w, if $\langle w, v \rangle > 0$ then w is on the positive half-space, and the sign is changed, than w is on the negative half-space. The set $\{w \in \mathbb{R}^n : \langle w, v \rangle = 0\}$ is the hyperplane orthogonal to v. Make sure you

[4] You might want to look up the definition of epigraph, and the characterization of convex functions via their epigraphs, as well as the definition of quasi-convex functions. This is not necessary to understand these notes.

can visualize the hyperplane, which side is the positive one, and which side is the negative one.

You can readily find definitions and illustrations of these concepts on Wikipedia or in the references listed in the end.

Why Convex Problems Are So Useful

The good thing about convex optimization problems is that a local maximum is always a global maximum. It follows that any feasible points that satisfy the first-order conditions of the problem are maximizers of the function. In particular, we do not have to check second-order conditions. Additionally, if the function is strictly convex and a maximizer exists, then the maximizer is unique. A sufficient condition to guarantee existence of a maximizer is that the function be continuous, and that the feasible set be compact (that is known as Weierstrass' Theorem).

Note all this is true if we substitute "maximization" for 'minimization", "maximizer" for "minimizer" and "concave function" for "convex function". The domain stays convex, there is no such thing as a concave set!

Convex optimization problems are also interesting from a practical point of view because there are algorithms that can solve these problems efficiently and reliably, and these algorithms are implemented in a variety of free and commercial solvers. This is in stark constrast with general nonconvex problems, where we either have to be content with a "good" local maximum, or with solving small-sized problems (few hundreds of variables, instead of thousands or even millions of variables for convex problems).[5]

Visual Intuition: One Equality Constraint

Figure 1 illustrates the maximizer of a function $f : \mathbb{R}^2 \to \mathbb{R}$ constrained to a single affine equality $h : \mathbb{R}^2 \to \mathbb{R}$. You might recognize this figure if you have taken a basic economics course.

[5] When we care about computing the solution, there are many useful subcategories of convex problems, like linear programs, geometric programs and semidefinite programs, and for each of these subcategories there are algorithms that take advantage of the special structure at hand. But note from our discussion above that even without a specialized solver, you should be good to go! All you need to make sure is that the routine you use to optimize your functions verifies that a point satisfies the first-order conditions before returning it as a solution. Nelder-Mead, Genetic Algorithms, Simulated Annealing algorithms do *not* have this property, and are also comparatively slow for convex problems; those algorithms are indicated for non-convex problems. In short, for convex problems, if your problem is smooth, try a Newton-type solver; if it is nonsmooth, try a cone solver. For more information, check the references in the end.

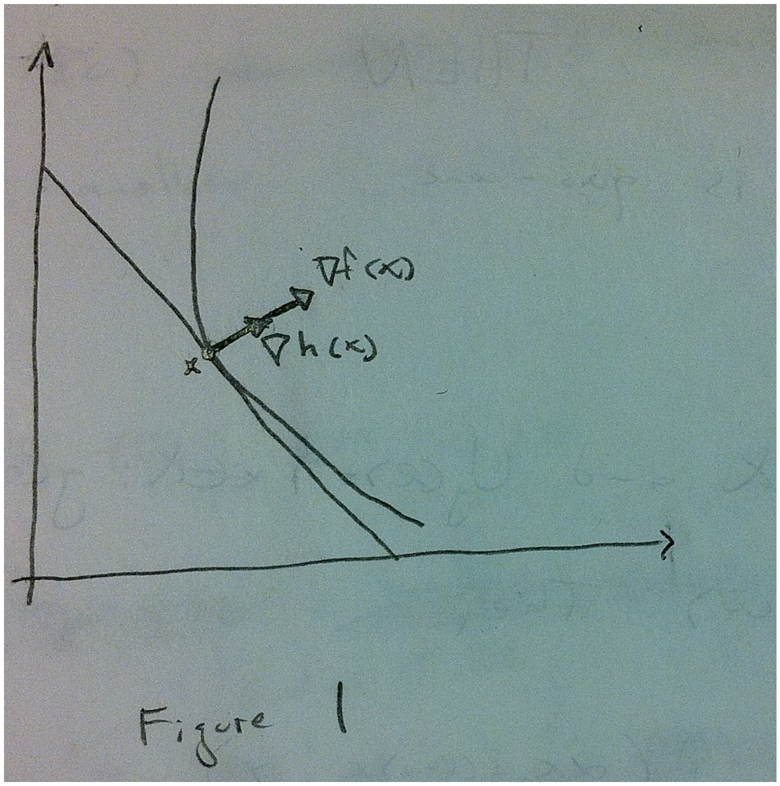

Figure 1 One equality constraint.

In this picture, it is clear that at a maximum, the level curve of the objective function f that goes through x has to be *tangent* to the affine constraint. Remembering that gradients are orthogonal to the level curves of their functions, it follows that the gradient of the objective function must be a *multiple* of the gradient of the constraint. With multiple constraints, what we obtain instead is that the gradient of the objective function must be a *linear combination* of the gradient of the constraints. The intuition for this is clear if we examine the case of multiple inequality constraints and realize that an equality constraint can be viewed as two inequality constraints.

Visual Intuition: Multiple Inequality Constraints

Figure 2 illustrates the maximizer of a function $f : \mathbb{R}^2 \to \mathbb{R}$ constrained to two convex inequalities $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$.

Intuitively speaking, it is clear that moving from a solution x as depicted above, we should move either in a direction that is a a direction of decrease for the objective function, or in a direction that is not feasible. In other words, "every direction of increase must be infeasible". Indeed, if we could move in a direction

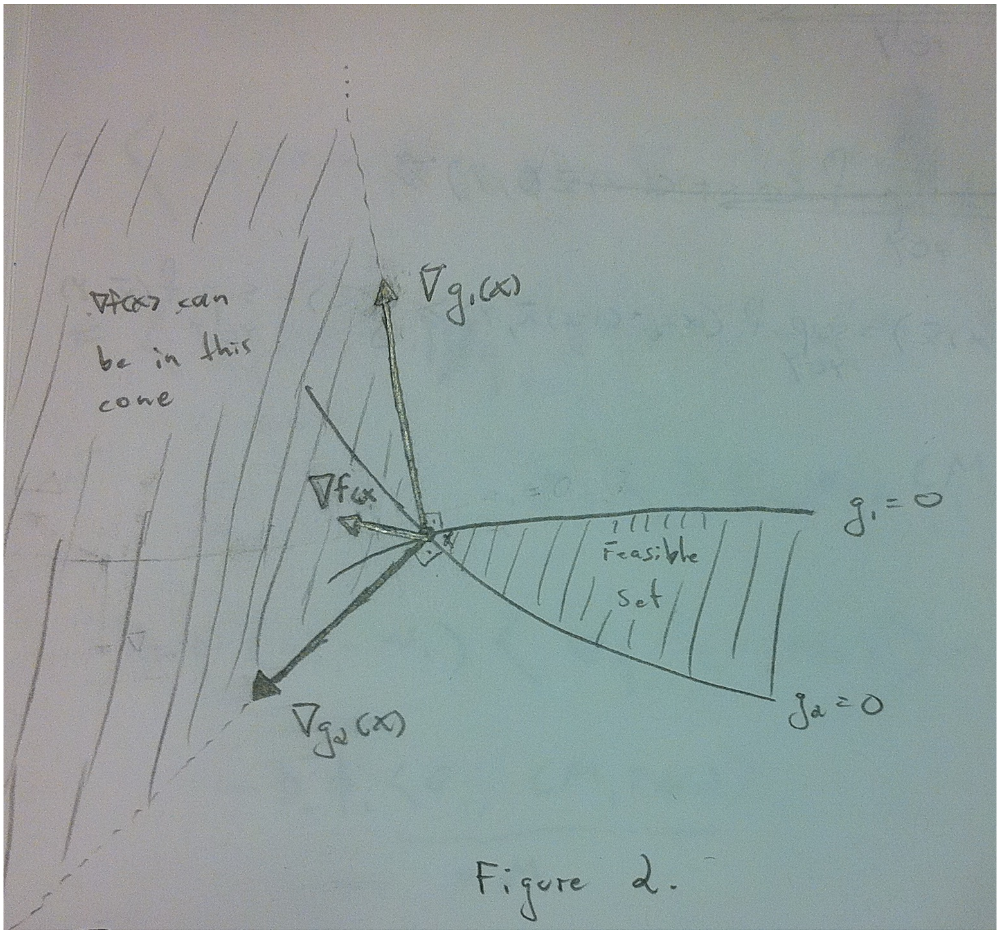

Figure 2 One equality constraint.

that was both feasible and ascending, we would obtain a feasible point with higher value than x, and thus x could not be a maximizer.

It turns out that given our smoothness and convexity assumptions, we can equate "direction of increase" with "direction that makes an acute angle with the gradient", and thus we can rewrite the statement "every direction of increase must be infeasible" as *there exists* $\mu_1 \geqslant 0$ *and* $\mu_2 \geqslant 0$ *such that*

$$\bigtriangledown f(x) = \mu_1 g_1(x) + \mu_2 g_2(x)$$

Take as long as you need to picture that.

Penalty Intuition: Approximating the Constraints with a Linear Function

The idea here is that with the right multipliers (that is, those that satisfy the KKT conditions), the Lagrangian is an auxiliary function that penalizes constraint violations linearly (and give a harmless "bonus" to strict feasibility). See Boyd and Vandenberghe's book, page 216, section 5.1.4.

Related to this, a common question I hear is: "I never know if I should write the Lagrangean as"

$$f(x) - \sum_{l=1}^{m} \lambda_l \, h_l(x) - \sum_{j=1}^{k} \mu_j \, g_j(x)$$

or as

$$f(x) + \sum_{l=1}^{m} \lambda_l \, h_l(x) + \sum_{j=1}^{k} \mu_j \, g_j(x)$$

The answer is that it depends on the sign of the inequality constraints and the multipliers. Remember, the Lagrangian has the intepretation of an auxiliary function that penalizes violations. To that end, if you have inequality constraints of the type $g(x) \leqslant 0$, then to penalize violations in the first Lagrangian above you will want $\mu \geqslant 0$, while in the second Lagrangian we need $\mu \leqslant 0$. The case of constraints $g(x) \geqslant 0$ is symmetric. If in doubt, multiply the constraints by -1 and apply the rule above.

Intuition for the Proof

The keys to making this intuition workable are the convexity and smoothness assumptions: they will allow us to rewrite the statement "every direction has to be either a direction of decrease or an infeasible direction" in the form of "either these linear inequalities hold, or these other linear inequalities hold, but not both", which is something known as a ***theorem of the alternative***.[6]

For the geometric intuition of the proof, we need to recall an important fact about continuously differentiable functions: any direction that is positive with respect to the gradient (that is, any direction d such that $\langle d, \bigtriangledown f(x) \rangle > 0$ is a direction of increase of f at x (note that f here is any function continuously differentiable at x, not only our objective function). It follows that if d is a direction of increase for f at x, then $\langle d, \bigtriangledown f(x), \geqslant \rangle 0$. The optimality condition (look at the figure for

[6] The one usually employed here is Farka's Lemma.

the inequality constrained case) is then that the at a solution x all directions of increase must be infeasible[7]. That translates to the condition that the gradient must be in the positive half-space of *all* binding inequality constraints. Check the figure for the inequality constrained case before and try to convince yourself that this is the case.

Main Result

Before we state the main result formally, let us be precise about the regularity conditions the constraints need to satisfy. We say the constraints in the basic problem satisfy Slater's constraint qualification condition, or simply ***Slater's condition***, if there exists x in the relative interior[8] of C such that $h(x) = 0$ and for all $j \in \{1, \ldots, k\}$, we have $g_j(x) < 0$.

The main theorem is the following.

Theorem 1 (Karush-Kuhn-Tucker, Differentiable Convex Case)

Suppose Slater's condition holds in the basic problem. Then $x \in C$ is a solution if and only if there are real numbers $\lambda_1, \ldots, \lambda_m$ (one for each equality constraint) and *nonnegative* real numbers $\mu_1, \ldots, \mu_k$ (one for each inequality constraint) such that x is a solution to the following auxiliary problem.

$$\underset{x \in C}{\texttt{maximize}} \qquad f(x) - \sum_{l=1}^{m} \lambda_l \; h_l(x) - \sum_{j=1}^{k} \mu_j \; g_j(x)$$

The objective function of the auxiliary problem is called the ***Lagrangian*** and each one of the λ_l and μ_j is called the ***Lagrange multiplier*** of the corresponding constraint.

From the theorem above we obtain a more workable result.

Corollary 1

Suppose Slater's condition holds in the basic problem. Then $x \in C$ is a solution if and only if there are $\lambda \in \mathbb{R}^m$, and $\mu \in \mathbb{R}^k$ such that the following holds:

---

7 Actually, we need to look at a larger set of directions, called normal directions. But given that we are not giving precise definitions, infeasible is more intuitive. See the references in the end for more details.

8 If you are not familiar with that concept, use the stronger concept of *interior*.

$$
\begin{aligned}
\triangledown f(x) &= \sum_{l=1}^{m} \lambda_l \ \triangledown h_l(x) + \sum_{j=1}^{k} \mu_j \ \triangledown g_j(x) \\
h_l(x) &= 0, \quad \forall i \in \{1, \ldots, m\} \\
g_j(x) &\leqslant 0, \quad \forall j \in \{1, \ldots, k\} \\
\mu_j &\geqslant 0 \quad \forall j \in \{1, \ldots, k\} \\
\mu_j g_j(x) &= 0 \quad \forall j \in \{1, \ldots, k\}
\end{aligned}
$$

The system above is called the system of Karush-Kuhn-Tucker conditions of the basic problem, or simply the ***KKT system***.

The KKT system spells out the ***first-order conditions*** (FOC) for our basic problem. In short, the corollary above says that for convex problems satisfying Slater's condition, the FOC given by the KKT system are necessary and sufficient for a solution. A few comments are in order.

- Note that we can use the KKT system to *find* solutions to the basic problem. Indeed, seeing x, λ and μ as an unknowns, the KKT system has $n + m + k$ equations and the same number of unknowns. This is very useful when we want to analyze the basic problem. That said, when solving problems numerically, it is better to use optimization algorithms to solve the basic problem than to use equation-solving algorithms to solve the KKT system.

- Note that there are no guarantees that the Lagrange multipliers will be unique. A sufficient condition for this is that the gradients of the active constraints form a linearly independent set. This condition is called the *linear independency constraint qualification* (LICQ).

- In well behaved problems, the set of all possible Lagrange multipliers at a given solution should be at least compact — that is, closed and bounded. Look up the *Mangasarian-Fromovitz constraint qualification condition*. This is something useful to know because if you are solving some optimization problem in a computer and while the algorithm is searching for a solution the multipliers are getting larger and larger (not all algorithms try to estimate both the solution and the multipliers at the same time, but many do), then you should suspect that your problem is ill-behaved in the sense that the Mangasarian-Fromovitz condition does not hold. In that case, you should reexamine your problem and try to either remodel it or verify that the algorithm you are employing is robust to failures of the Mangasarian-Fromovitz condition.

- Similar results hold for nonsmooth convex problems, and locally, for smooth nonconvex problems. See references in the end. Also, even in the case above, we only need the functions to be continuously differentiable at the solution.

- For nonconvex problems, the KKT conditions are necessary, but not sufficient. That is not surprising if we remember that in general, for nonconvex problems we need to check second-order conditions, and the KKT conditions are only the first-order conditions of the problem.

A Word on Why Convex Problems are "Easy" to Solve

Roughly speaking, the reason why convex problems are easy to solve is that all we only need to know how the function behaves in a neighborhood of our current guess to take a step that makes a substantial improvement. An algorithm that is based on this idea of making a "locally optimal" choice at each step is what computer scientists call a ***greedy algorithm***. So, as long as your problem is convex, the "Gordon Gekko motto"[9] applies: "Greed is right. Greed works."

For a rigorous analysis of the complexity of convex optimization problems, check Nesterov and Nemirovski's work on self-concordant functions (there is a SIAM book they wrote on this). For a more accessible treatment, check Boyd and Vandenberghe's book "Convex Optimization".

Making Nonconvex Problems Convex

Now that we know that we can make rich analyses and easily solve convex optimization problems, we would like to turn as many problems as we can into convex problems. We present two simple tricks to achieve just that.

First, suppose you have an optimization problem where you have an equality constraint $g(x) = 0$, where $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function that is not affine. This constraint makes the problem nonconvex. One way to solve this problem is to relax the constraint by only requiring that $g(x) \leqslant 0$. By doing so, we recover convexity, but by enlarging the feasible set, we might find a solution that is not feasible in the original problem. There are two typical ways to deal with this problem; you can either prove that in any solution to the relaxed problem we will have $g(x) = 0$ (that is, the constraint will bind) or you can try to solve the relaxed problem and verify afterwards if it satisfies the original constraint $g(x) = 0$. Even if it doesn't, you might learn something useful in this exercise. These techniques are known as ***convex relaxation*** because we restored convexity by relaxing a constraint.[10]

---

[9] Character played by Michael Douglas in the movie *Wall Street*.

[10] Note that this showed up in the problem set 1, problem 2, where the max-surplus problem was a problem of maximizing a concave function subject to an equality constraint given by a convex function.

Now suppose you have a maximization problem where the objective function is given by $f(x) = \min\{r_i(x) : i \in I\}$, where I is some index set and each r_i is concave and differentiable. The function f is still concave, but it is certainly not differentiable in general. There is a simple trick to turn this into a differentiable convex optimization problem. First, we add an auxiliary variable t and we maximize a new objective function $\tilde{f}(x,t) = t$ subject to the original constraints and new constraints $t \leqslant f(x)$. This trick is called the ***epigraph transformation***. However, we now we have not made much progress, as now we have a nondifferentiable constraint $f(x)$. This new hurdle, however, is easy to bypass if we note that $t \leqslant f((x)$ if and only if $t \leqslant r_i(x)$ for all $i \in I$. By adding one constraint for each one of the functions r_i, we have restored convexity to our problem! To understand the name of the trick, look up the definition of epigraph; essentially, we are moving a problem from the domain of the objective function to its epigraph.

For more details on these techniques and more, check Boyd and Vandenberghe's book *Convex Optimization*.

Epilogue

What did we leave out?

- Duality. Very rich theory with many applications.
- The envelope theorem and Lagrange multipliers as sensitivities/shadow prices.
- Many other things that are useful only for specific problems.

Notes on References

(Note to myself: Reorganize it later in a proper References list)

There are many more references on this subject than I will list here, so I will only mention references that are in some sense close to these notes. The book *Convex Optimization* by Boyd and Vandenberghe is superb, and freely available on the internet, just google "convex optimization". The chapter on Lagrange duality is especially relevant to this material. The main theme of the book is "how to reformulate almost everything as a convex optimization problem", with many techniques beyond convex relaxation, and they are the only book that I know that do that. They have a good mix of theory, modeling and algorithms. They also provide free code for convex optimization problems and pointers to good commercial code. Google the book and you should also find some very good lectures given by Stephen Boyd. Kim Border has many notes on the subject on his website, the notes for Ec 181 should be particularly useful. Truman Bewley has excellent notes on the convex but nondifferentiable case. They are very short and rigorous, and should be available at `http://www.econ.yale.edu/graduate/mathcamp/lec09.pdf`. If they are not, let me know, as I have a copy of them and I can forward them to you. Finally, Nocedal and Wright wrote a widely praised book called *Numerical Optimization* that has an excellent chapter on the theory of constrained optimization, with very good pictures. I'll provide you guys with some solved examples as soon as I can.