4. More advanced topics in optimization

Note: This lecture owes most of its material to the book by Boyd and Vandenberghe (2009).

4.1. **Equivalent (and simpler) problems.** Recall that last time, we defined a notion of equivalence for optimization problems: informally, two problems are equivalent if the solution to one can be immediately obtained from the solution to the other, and vice versa. It's important to keep a handful of transformations in mind that lead to equivalent (but potentially simpler) problems.

As our first example, consider a strictly convex quadratic objective, with some of the variables unconstrained:

$$\begin{aligned} \text{minimize } & f_0(x_1, x_2) := x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2 \\ \text{subject to } & f_i(x_1) \le 0, i = 1, \dots, m, \end{aligned}$$

where P_{11} and P_{22} are symmetric. Here we can analytically minimize over x_2:

$$\inf_{x_2} f_0(x_1, x_2) = x_1^T \left(P_{11} - P_{12} P_{22}^{-1} P_{12}^T\right) x_1$$

More generally, we always have

$$\inf_{x,y} f(x,y) = \inf_x \tilde{f}(x) \quad \text{where} \quad \tilde{f}(x) := \inf_y f(x,y).$$

In other words, we can always minimize a function by first minimizing over some of the variables, and then minimizing over the remaining ones.

As our second example, consider eliminating linear equality constraints, ie. those of the form $Ax = b$, which can always be done in any problem. Assuming the problem is feasible, let x_0 denote any solution of the equality constraints. Let $F \in \mathbb{R}^{n \times k}$ be any matrix with range$(F) =$ null(A), so the general solution of the linear equations $Ax = b$ is given by $Fz + x_0$, where $z \in \mathbb{R}^k$. We can choose F to be full rank, in which case we have $k = n - r$ where $r =$ rank(A). Substituting $x = Fz + x_0$ into the original problem yields the problem

$$\begin{aligned} \text{minimize } & f_0(Fz + x) \\ \text{subject to } & f_i(Fz + x_0) \le 0, \quad i = 1, \dots, m, \end{aligned}$$

with variable z, which is equivalent to the original problem, has no equality constraints, and r fewer variables. This can be done in reverse when the objective and constraint functions are given as compositions of the functions f_i with affine transformations $A_i x + b_i$.

Recall from last time that the *optimal value* of an optimization problem (not necessarily convex) will be denoted

$$p^* = \inf\left\{f_0(x) \mid f_i(x) \le 0,\ h_j(x) = 0\ (\forall i, j)\right\}$$

By convention, $p^* = \infty$ if the problem is infeasible (no x satisfies the constraints), and $p^* = -\infty$ if the problem is unbounded below.

Recall also the **Lagrange dual problem**:

$$\text{maximize } g(\lambda, \nu) \text{ subject to } \lambda \succeq 0.$$

If we could solve this problem, we would know the best lower bound on p^* that could be obtained by applying the lower bound property of the dual function. This is generally a convex optimization problem, even if the original one wasn't! The optimal value of the dual problem is typically denoted d^*.

Equivalent formulations of a problem can lead to very different duals. Reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting. For example, consider the problem

$$\text{minimize } f_0(Ax + b)$$

The dual function is constant:

$$g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^*,$$

and finding that constant is simply solving the original problem. However, suppose we write it as

$$\text{minimize } f_0(y) \ \text{ subject to } Ax + b - y = 0.$$

Then the dual problem is

$$\text{maximize}(b^T \nu - f_0^*(\nu)) \ \text{ subject to } A^T \nu = 0.$$

4.2. **General duality gaps.** The statement $d^* \le p^*$ always holds, and is known as *weak duality*. It is sometimes useful to find nontrivial lower bounds for difficult problems. The equality $d^* = p^*$ does not usually hold. When it holds, it is known as *strong duality*. It usually does hold for convex problems, under conditions that we'll make precise very soon. The difference $p^* - d^* \ge 0$ is known as the *duality gap*.

More generally, dual feasible points allow us to bound how suboptimal a given feasible point is, without knowing the exact value of p^*. Indeed, if x is primal feasible and (λ, ν) is dual feasible, then

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu).$$

In particular, this establishes that x is ϵ-suboptimal, with $\epsilon = f_0(x) - g(\lambda, \nu)$.

We refer to the gap between primal and dual objectives, $f_0(x) - g(\lambda, \nu)$, as the duality gap associated with the primal feasible point x and dual feasible point (λ, ν). A primal-dual feasible pair $x, (\lambda, \nu)$ localizes the optimal value of the primal (and dual) problems to an interval:

$$p^*, d^* \in [g(\lambda, \nu), f_0(x)],$$

the width of which is the duality gap.

4.3. **Slater's condition.** Any condition that guarantees strong duality in convex problems is called a *constraint qualification*. The most well-known one is Slater's constraint qualification. To state Slater's condition, we need to define the **affine hull** $\mathrm{aff}(S)$ of a set S, which is the smallest affine set containing S, or equivalently, the intersection of all affine sets containing S. Recall that in any normed vector space, we have a notion of distance, $d(x, y) = \|x - y\|$, giving rise to a notion of an open ball, $B(x, \epsilon) = \{y : d(x, y) < \epsilon\}$. The **relative interior** of a set S is defined as

$$\mathrm{relint}(S) := \{x \in S : \exists \epsilon > 0, B(x, \epsilon) \cap \mathrm{aff}(S) \subseteq S\},$$

If this definition seems abstract, just think of the set $\{(x, y, 0) \in \mathbb{R}^3 \mid x^2 + y^2 \leq 1\}$. Its interior is empty, but its relative interior is an open unit disk in the xy-plane.

Theorem 4.1 (Slater's condition)**.** *Strong duality holds for a convex problem if it is* strictly feasible*, ie*

$$\exists\, x \in \mathrm{relint}(\mathcal{D}) \;:\; f_i(x) < 0 \text{ for } i = 1, \ldots, m, \text{ and } Ax = b. \tag{4.1}$$

Under these conditions, the dual optimum is attained if $p^* > -\infty$.

The theorem can be sharpened in various ways. For example, affine inequalities do not need to hold with strict inequality. Note that there exist non-convex problems with strong duality, it just isn't guaranteed in general. Proving Theorem 4.1 is outside of our current scope, but if we have time later we may return to the methods which would allow us to prove it. The proof isn't just a few lines.

4.4. **Karush-Kuhn-Tucker conditions.** We'll now begin some considerations which will lead us up to the important *Karush-Kuhn-Tucker (KKT) conditions*, which are practically what one uses to actually solve many optimization problems. Assume strong duality holds, x^* is primal optimal, (λ^*, ν^*) is dual optimal. Then

$$\begin{aligned} f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_j \nu_j^* h_j(x) \right) \\ &\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_j \nu_j^* h_j(x^*) \\ &= f_0(x^*) \end{aligned}$$

To get from the second line to the third, note that all $h_j(x^*) = 0$, and $\lambda_i^* f_i(x^*) \leq 0$ by our conventions and definitions, but if it were strictly negative, then we'd have $f_0(x^*) = f_0(x^*) + \text{negative}$, which is a contradiction. Hence the inequality is actually an equality.

This gives us a way of seeing which constraints are active at the solution. Since all of the terms of the form $\lambda_i^* f_i(x^*)$ must vanish, if $\lambda_i^* > 0$ then $f_i(x^*)$ must be zero, meaning the constraint is active, or alternatively if $f_i(x^*) < 0$ (meaning the constraint isn't active) then $\lambda_i^* = 0$. So any constraints which are barely satisfied (or "active" or which "are binding") will have nonzero Lagrange multipliers in the solution to the dual problem. Any non-binding constraints will have their corresponding Lagrange multipliers in the dual-optimal solution equal to zero. The above chain of inequalities also directly imply that x^* minimizes $L(x, \lambda^*, \nu^*)$.

Definition 4.1. The following four conditions are called the **KKT conditions** (for a problem with differentiable f_i, h_i):

(1) primal constraints: $f_i(x) \leq 0$, $i = 1, \dots, m$, $h_i(x) = 0$, $i = 1, ..., p$
(2) dual constraints: $\lambda \succeq 0$
(3) complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \dots, m$
(4) The gradient of the Lagrangian with respect to x vanishes:

$$\nabla_x f_0(x) + \sum_i \lambda_i \nabla f_i(x) + \sum_j \nu_j \nabla h_j(x) = 0$$

From our discussion above, if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions. This is the case for convex and non-convex problems alike.

Furthermore, if $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal. To see this, note that the first two KKT conditions state that $\tilde{x}$ is primal feasible. Since $\tilde{\lambda} \succeq 0$, it follows that $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x; the last KKT condition states that its gradient with respect to x vanishes at $x = \tilde{x}$, so it follows that $\tilde{x}$, minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ by the first-order condition. Furthermore,

$$f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = g(\tilde{\lambda}, \tilde{\nu}).$$

In summary, for any convex optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

If Slater's condition is satisfied in a convex problem, then x is optimal if and only if there exist λ, ν that satisfy the KKT conditions with the given x. This can be viewed as a generalization of the optimality condition $\nabla f_0(x) = 0$ for unconstrained problems.

4.5. **Quadratic programming.** An important example in which the KKT conditions can be written down, and solved, explicitly is *equality constrained convex quadratic minimization*. We consider the problem

$$\text{minimize } (1/2)x^T P x + q^T x + r$$
$$\text{subject to } Ax = b,$$

where $P \in S^n_+$. The KKT conditions for this problem are

$$Ax^* = b, \quad Px^* + q + A^T \nu^* = 0$$

which we can write equivalently as a single matrix equation,

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \nu^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

Solving this set of $m + n$ equations in the $m + n$ variables (x^*, ν^*) gives the optimal primal and dual variables. Assuming the relevant inverses exist, one could apply the block inversion formula to get an explicit formula for x^* that doesn't depend on ν^* (exercise).

An important financial application of the above occurs when we have a factor model for asset returns:

$$r_{t,t+1} = X_t f_{t,t+1} + \epsilon_{t,t+1}$$

in which the matrices X_t are entirely composed of risk factors, ie. factors that are believed to help forecast risk, as opposed to alpha factors, or factors that are believed to help forecast return. Indeed, assume our aggregate forecast α_t is orthogonal to the columns of X_t. A subscript of t on any variable means that the associated value could have been calculated as of time t. The variance-covariance matrix in this model is then

$$V_t = \mathrm{cov}(r_{t,t+1}) = X_t F_t X_t' + \Delta_t,$$

where

$$F_t := \mathrm{cov}(f_{t,t+1}) \in S_+^k, \quad \Delta_t := \mathrm{diag}(\mathrm{var}(\epsilon^1), \ldots, \mathrm{var}(\epsilon^n)).$$

Consider unconstrained mean-variance optimization:

$$\text{minimize } \ f_0(x) := \frac{\gamma}{2} x' V_t x - \alpha_t \cdot x$$

where $\gamma > 0$ is the absolute risk aversion index of Arrow (1971) and Pratt (1964).

Note that under the assumptions above, in particular the assumption that $\alpha_t' X_t = 0$, we can always reduce factor risk (without changing the expected return of the portfolio) by trading until our factor exposures are zero. We might increase idio risk by a small amount in the process, but this can be neglected if an approximation is desired. So x^* will, approximately, lie near the space orthogonal to the columns of X_t, ie. $x^* \in X_t^{\perp}$.

Even if $\alpha_t' X_t \neq 0$, we can always solve the mean-variance problem with equality constraints:

$$\begin{aligned} \text{minimize } \ & \frac{\gamma}{2} x' V_t x - \alpha_t \cdot x \\ \text{subject to } \ & X_t' x = 0\,. \end{aligned}$$

Note that for any feasible x, one has $x' V_t x = x' \Delta_t x$ so the quadratic part of the problem only depends on inverting a *diagonal* matrix. In your empirical work, if you ever find yourself inverting an $n \times n$ covariance matrix where n is the number of assets, ask yourself whether there is a factor model for the assets you're trading. If there is, you probably shouldn't be inverting an $n \times n$ matrix. If $\alpha_t' X_t \neq 0$ then the equality-constrained solution will be suboptimal relative to the unconstrained solution, but it's only "missing the alpha" that is aligned with the risk factors, which arguably shouldn't be there anyway, and it has the benefit that the only matrix being inverted is diagonal!

Now consider a version of the above in which we have linear inequality constraints. For simplicity, we'll drop the linear term in f_0. So the "primal" or original problem is, for $P \in S_{++}^n$,

$$\text{minimize } x^T P x \ \text{ subject to } Ax \preceq b.$$

The dual function is

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda \,.$$

The dual problem is then maximizing a quadratic-plus-linear function $g(\lambda)$ over the space $\lambda \succeq 0$, which seems simpler. Slater's condition specifies that $p^* = d^*$ if $A\tilde{x} \prec b$ for some $\tilde{x}$, but in fact more is true: for this class of problems, $p^* = d^*$ in any case.

References

Arrow, Kenneth J (1971). "Essays in the theory of risk-bearing". In:

Boyd, Stephen and Lieven Vandenberghe (2009). *Convex optimization*. Cambridge university press.

Pratt, John W (1964). "Risk aversion in the small and in the large". In: *Econometrica: Journal of the Econometric Society*, pp. 122–136.