1. Convex Sets

(1) The set of optima for a min LP is a convex set (just use linearity of the matrix defining the constraints, i.e. since $Ax \preceq b$ defines the polyhedron).
(2) Intersection, Minkowski sum, Minkowski difference, preserve convexity of sets.
(3) Caratheodory's theorem: for a convex set in $\mathbb{R}^n$ any point can be expressed as a convex combination of at most $n+1$ points (prove using linear independence of differences between points).
(4) For S a convex set with non-empty interior, x_1 in the closure of S and x_2 in the interior then the "weak" convex combination, i.e. ($\lambda \in (0,1)$) is always in the interior.
(5) The interior of a convex set is convex and if the interior is non-empty then so is the closure.
(6) For a convex set with non-empty interior $cl\,(int\,(S)) = cl\,(S)$ and $int\,(cl\,(S)) = int\,(S)$.

2. Hyperplanes

2.1. **Projection theorem.** Projection theorem: for any closed convex set S and for any point not in the convex set y there exists a projection $\bar{x}$ of y onto S and $\bar{x}$ is such that $(y-\bar{x})^{\intercal}(x-\bar{x}) \leq 0$ for all $x \in S$, i.e. the plane defined by $(y-\bar{x})$ separates S from y.

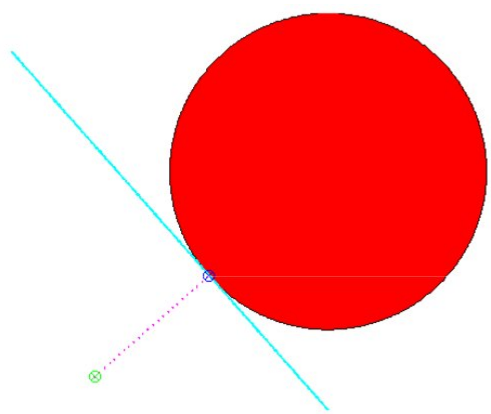

Figure 2.1. Projection theorem

Proof. We seek to minimize $\|x-y\|$ over $x \in S$. This is equivalent to minimizing over the set

$$\left\{x \middle| \, \|x-y\| \leq \|x-w\|\right\}$$

for some $w \in S$. Why? Because $\left\{x \middle| \, \|x-y\| \leq \|x-w\|\right\} \subset S$ and the $\bar{x}$ that minimizes $\|x-y\|$ is definitely in it (since it's smaller than all $\|x-y\|$). This is a bounded closed set (since S is closed) and norms are continuous and so by extreme value theorem there exists a minimum. $\bar{x}$ is unique: suppose there exist two such points $\bar{x}_1, \bar{x}_2$. Then

$$\begin{aligned} 0 &< \|(\bar{x}_1-y)-(\bar{x}_2-y)\|^2 \\ &= 2\|\bar{x}_1-y\|^2 + 2\|\bar{x}_2-y\|^2 - 4\left\|\frac{1}{2}\left[(\bar{x}_1-y)-(\bar{x}_2-y)\right]\right\|^2 \\ &= 2\|\bar{x}_1-y\|^2 + 2\|\bar{x}_2-y\|^2 - 4\left\|\frac{(\bar{x}_1+\bar{x}_2)}{2} - y\right\|^2 \\ &= 2\|\bar{x}_1-y\|^2 + 2\|\bar{x}_2-y\|^2 - 4\|\hat{x}-y\|^2 \end{aligned}$$

Since $\|\bar{x}_1-y\| = \|\bar{x}_2-y\| = s$ rearranging we have $\|\hat{x}-y\| < s$ contradicting that x_1, x_2 are minimal projections.

Corollary. *Let S be the column space of some matrix A and $y \notin S$. Then for the projection $\bar{z}$ it's the case that $(y - \bar{z}) \perp z$ for all $z \in col(A)$ and $A^{\intercal}\bar{z} = A^{\intercal}y$ (i.e. the projection is the perpendicular projector).*

Corollary. *Least squares.*

$\square$

Improper separation is when possibly both sets are in the the separating hyperplane. **Proper separation** is when the union of the two sets isn't contained in the separating hyperplane. **Strict separation** is when the convex sets don't intersect the separating hyperplane (but their closures might). **Strong separation** is when there's "fat" in between the sets and the hyperplane.

Point-to-set separation: for any non-empty closed convex set S and $y \notin S$ there's a separating hyperplane. Proof: use the projection theorem to find the projection. Keep in mind the obtuseness $((y - \bar{x})^{\intercal}(x - \bar{x}) \leq 0$ says that every vector from $x \in S$ to $\bar{x}$ is obtuse to the normal vector to the separating hyperplane $p = y - \bar{x})$.

Corollary. *Every closed convex set is the intersection of halfspaces (take all the separation hyperplanes and intersection the half spaces that the set is in).*

Corollary. *Let S be nonempty and y not in the closure of the convex hull of S. Then you can stronly separate.*

2.2. **LP Duality.**

Theorem. *Farkas' lemma: exactly one of the two following systems has a solution*

$$x \text{ s.t. } Ax \preceq 0 \text{ and } c^{\intercal}x > 0$$
$$y \text{ s.t. } A^{\intercal}y = c \text{ and } y \succeq 0$$

*The intuition here is if the columns of $A^{\intercal}$ by $a_1, \ldots, a_m$ the second system has a solution iff c lies in the conic cone??? of $a_1, \ldots, a_m$. Apparently a convex cone is a cone closed under conic combinations and a cone is any set closed under positive scalings. If c doesn't lie in the cone then the polyhedral convex cone[1] $Ax \preceq 0$ and the halfspace $c^{\intercal}x > 0$ have a nonempty intersection. Note that cone defined by this system is actually the polar of the vectors themselves (since these vectors actually define the **normals** to the plane). This is basically all about polar cones. Either a vector is in the convex cone or there exists a vector in the polar cone that makes an accute angle with it (though not necessarily in the polar cone).*

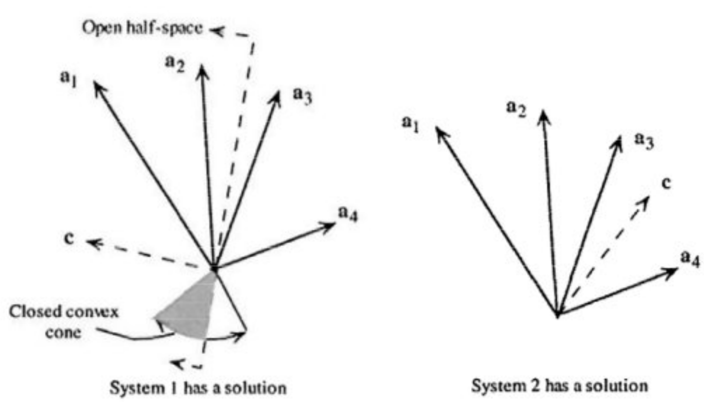

FIGURE 2.2. Farkas lemma

[1]Why is this a polyhedral cone? Well x that satisfies $Ax \preceq b$ is the intersection of half-spaces (think hyperplanes defined by the rows of A). If $b = 0$ then all of those halfspaces go through the origin.

Proof. Suppose system 2 has a solution, i.e. c is in the polar cone of the rows of A. Then there exists $y \succeq 0$ such that $A^\intercal y = c$. Let x be in the other cone, i.e. $Ax \preceq 0$. Then $c^\intercal x = y^\intercal Ax \preceq 0$ (since $y \succeq 0$ times $Ax \preceq 0$ is $\preceq 0$). Hence system 1 has no solution. Suppose system 2 doesn't have a solution. I.e. c doesn't lie in the convex cone of the rows of A. Let $S = \{x | x = A^\intercal y, y \succeq 0\}$ i.e. the convex cone. S is closed convex set and $c \notin S$. By point set separation there exists a vector p and scalar α such that $p^\intercal c > \alpha$ and $p^\intercal x \leq \alpha$ for all $x \in S$ (p is the projection vector from S to c and all this says is that there's a plane defined by the normal vector p that separates c and S with S on the "negative" side of that plane. Since $0 \in S$ (since we're talking about a convex cone, $y = 0$) $\alpha \geq 0$ (just take one of the faces of S) so $p^\intercal c > 0$. Also $\alpha \geq p^\intercal A^\intercal y = y^\intercal Ap$ (because $x = A^\intercal y$ and $\alpha \geq p^\intercal x$) for all $y \succeq 0$. Since y can be made arbitrarily large the last inequality implies that's really negative is Ap, i.e. $Ap \preceq 0$. Therefore p is a vector such that $Ap \preceq 0$ and $p^\intercal c = c^\intercal p < 0$. □

Corollary. *Gordan's theorem*

$$\{x | Ax \prec 0\} = \emptyset$$
$$\iff$$
$$\{y | A^\intercal y = 0, y \succ 0\} = \emptyset$$

Proof. Note $Ax \prec 0$ iff $A\mathbf{x} + \mathbf{e}s \preceq 0$ which is equivalent to

$$[A\,\mathbf{e}]\begin{bmatrix}\mathbf{x}\\ s\end{bmatrix} \preceq 0 \quad (0, \ldots, 0, 1)\begin{bmatrix}\mathbf{x}\\ s\end{bmatrix} \succ 0$$

By Farkas' lemma the associated system is

$$\begin{bmatrix}A^\intercal\\ \mathbf{e}^\intercal\end{bmatrix} y = (0, \ldots, 0, 1)\ y \succeq 0$$

I.e. $A^\intercal y = 0$ and $\mathbf{e}^\intercal y = 1$. So $y \neq 0$. By Farkas' lemma these two systems are alternative systems. □

Corollary. *Another form of Farkas' lemma*

$$\{x | Ax \succeq b\} = \emptyset$$
$$\iff$$
$$\{y | A^\intercal y = 0, b^\intercal y \succ 0\} = \emptyset$$

2.3. **Duality of LPs.** Consider the primary system

$$\begin{aligned}&\max 2x_1 + 3x_2\\ &\text{s.t} 4x_1 + 8x_2 \leq 12\\ &2x_1 + x_2 \leq 3\\ &3x_1 + 2x_2 \leq 4\\ &x_1, x_2 \geq 0\end{aligned}$$

Consider solving the problem by successively bounding it from above. Note that since

$$2x_1 + 3x_2 \leq 4x_1 + 8x_2 \leq 12$$

$\max(2x_1 + 3x_2) \leq 12$. Similarly

$$2x_1 + 3x_2 \leq \frac{1}{2}(4x_1 + 8x_2) \leq 6$$

or even more creatively

$$2x_1 + 3x_2 \leq \frac{1}{3}(4x_1 + 8x_2) + (2x_1 + x_2) \leq 5$$

So basically the game is equating coefficients. You can see that if we let y_1, y_2, y_3 be the linear combination coefficents and enforce that after equating the coeffients they don't fall below 2 for x_1 and 3 for x_2, i.e.

$$2x_1 + 3x_2 \leq y_1(4x_1 + 8x_2) + y_2(2x_1 + x_2) + y_3(3x_1 + 2x_2) \leq y_1 12 + y_2 3 + y_3 4$$

or

$$\begin{aligned}4y_1 + 2y_2 + 3y_3 &\geq 2\\ 8y_1 + y_2 + 2y_3 &\geq 3\\ y_1, y_2, y_3 &\geq 0\end{aligned}$$

and we minimize $12y_1 + 3y_2 + 4y_3$ we should get a tight upper bound on the object of the original problem. This is obviously an LP as well (called the dual). And you can go backwards as well. In general we have the **primal**

$$\begin{aligned}&\max_x c^\intercal x\\ &\text{s.t}\, Ax \preceq b\\ &x \succeq 0\end{aligned}$$

and the **dual**

$$\begin{aligned}&\min_y b^\intercal y\\ &\text{s.t}\, A^\intercal y \succeq c\\ &y \succeq 0\end{aligned}$$

and what holds is that either both problems are infeasible, infeasible and unbounded respectively (or vice versa) or both feasible and with optima equal. This is called **strong duality** in LPs. On the way to proving that we need to prove **weak duality**

Theorem. *Weak duality: for the primal solution x and dual problem solution y*

$$c^\intercal x \leq b^\intercal y$$

Proof. To wit

$$\begin{aligned}c^\intercal x &= x^\intercal c\\ &\leq x^\intercal \left(A^\intercal y\right) \text{ since } y \text{ is feasible for the dual and } x \succeq 0\\ &= (Ax)^\intercal y\\ &\leq b^\intercal y \text{ since } x \text{ is feasible for the primal and } y \succeq 0\end{aligned}$$

□

Theorem. *Strong duality for LPs. We have to rewrite the dual and primal a little: primal*

$$\begin{aligned}&\min_x c^\intercal x\\ &s.t\, Ax \succeq b\end{aligned}$$

and the dual

$$\begin{aligned}&\max_y b^\intercal y\\ &s.t\, A^\intercal y = c\\ &y \succeq 0\end{aligned}$$

Proof. Suppose the dual is feasible and its max is δ. Let

$$P' = \left\{x \middle| Ax \succeq b, c^\intercal x \leq \delta\right\}$$

If P' is nonempty then the primal must have a feasible solution with value at most δ (since P' is basically the reduced feasible region of the primal. Note that $P' = \left\{x \middle| Ax \succeq b, -c^\intercal x \geq -\delta\right\}$. Towards a contradiction suppose P' is empty. Then by Farkas' lemma (form 2) there exist y, λ such that

$$\begin{bmatrix}A^\intercal\\ -c\end{bmatrix}\begin{pmatrix}y\\ \lambda\end{pmatrix} = 0 \text{ and } (b, -\delta)^\intercal \begin{pmatrix}y\\ \lambda\end{pmatrix} > 0$$

This implies that $A^\intercal y - c\lambda = 0$ and $b^\intercal y - \lambda\delta > 0$. There are two cases

Case 1. If $\lambda = 0$ then $A^\intercal y = 0$ and $b^\intercal y > 0$. Choose $z \succeq 0$ such that $A^\intercal z = c$ and $b^\intercal z = \delta$. Then for $\varepsilon > 0$

$$\begin{aligned} A^\intercal (z + \varepsilon y) &= 0 \\ z + \varepsilon y &\succeq 0 \text{ since } y \succeq 0 \\ b^\intercal (z + \varepsilon y) &= \delta + \varepsilon b^\intercal y > \delta \end{aligned}$$

so $(z + \varepsilon y)$ is a feasible solution of the dual with value greater than δ, a contradiction.

Case 2. Otherwise scale y and λ such that $\lambda = 1$ (since both y, λ are nonegative) and so $A^\intercal y = c$ and $b^\intercal y > \delta$. This means y is a solution of the dual with value greater than δ, a contradiction.

Therefore P' is feasible, so the primal is feasible with value at most δ. By weak duality its value is at least δ. Hence the primal solution and the dual solution are equal. □

Another way to look at it is a generalization of Lagrange multipliers: consider the following constained minimization problem

$$\begin{aligned} &\max_x x^2 + y^2 \\ &\text{s.t. } x + y = 1 \end{aligned}$$

Let $L(x, y, \lambda) = x^2 + y^2 + \lambda(1 - x - y)$. Think of solving the original problem by, instead of enforcing $x + y = 1$, allow it to be violated and associate a cost $\lambda(1 - x - y)$, with cost rate λ. This is then an unconstrained minimization problem over x, y, λ: first minimize with respect to x, y

$$\nabla_{x,y} L = (2x - \lambda x, 2y - \lambda y) = 0$$

Solving for x, y we get that$x = y = \frac{p}{2}$. Then the constraint $x + y = 1$ gives the additional relation $p = 1$ and hence the optimal solution to the original problem is $x = y = 1/2$. Another way of interpreting this is: when the costrate is properly chosen ($p = 1$) the optimal solution to the constrained problem is also the optimal solution to the unconstrained problem.

For LPs consider the standard form problem

$$\begin{aligned} &\min_x c^\intercal x \\ &\text{s.t.} Ax = b \\ &x \succeq 0 \end{aligned}$$

called the primal problem. Assume an optimal x^* exists. A relaxed problem is

$$\begin{aligned} &\min_{x,\lambda} c^\intercal x + \lambda^\intercal (b - Ax) \\ &\text{s.t.} x \succeq 0 \end{aligned}$$

Why is this a relaxed problem? Well there are fewer constraints for one, and it turns out the objective always lower than the optimal value of the primal:

$$g(\lambda) = \min_{x \succeq 0} \left[c^\intercal x + \lambda^\intercal (b - Ax)\right] \leq c^\intercal x^* + \lambda^\intercal (b - Ax^*)$$

where the inequality follows because of the min. Then $c^\intercal x^* + \lambda^\intercal (b - Ax^*) = c^\intercal x^*$ since x^* is assumed to satisfy the original system. Therefore $g(\lambda)$ is always a lower bound for the primal and **maximizing** over λ then yields the tightest lower bound. What does this look like?

$$\begin{aligned} g(\lambda) &= \min_{x \succeq 0} \left[c^\intercal x + \lambda^\intercal (b - Ax)\right] \\ &= \lambda^\intercal b + \min_{x \succeq 0} (c^\intercal - \lambda^\intercal A) x \end{aligned}$$

Note that $\min_{x\succeq 0}\left(c^{\intercal}-\lambda^{\intercal}A\right)x=\min_{x\succeq 0}\left(c^{'}\right)^{\intercal}x$ is an LP over the positive orthant polyhedron. So if $\left(c^{\intercal}-\lambda^{\intercal}A\right)\succeq 0$ then the minimum is at $x=0$, otherwise if there exists a coordinate of $\left(c_i^{\intercal}-\lambda^{\intercal}A_i\right)<0$ then we can crank x_i arbitrarily large and the problem is unbounded. Hence

$$\min_{x\succeq 0}\left(c^{\intercal}-\lambda^{\intercal}A\right)x=\begin{cases}0 & \text{if } \left(c^{\intercal}-\lambda^{\intercal}A\right)\succeq 0\\ -\infty & \text{o/w}\end{cases}$$

Clearly maximizing $g(\lambda)$ can only happen when the inner minimization isn't equal to $-\infty$. So

$$\begin{aligned}&\max_{\lambda}\lambda^{\intercal}b\\ &\text{s.t.}\left(c^{\intercal}-\lambda^{\intercal}A\right)\succeq 0\end{aligned}$$

or

$$\begin{aligned}&\max_{\lambda}\lambda^{\intercal}b\\ &\text{s.t.}\lambda^{\intercal}A\preceq c^{\intercal}\\ &\lambda \text{ free}\end{aligned}$$

Compare with

$$\begin{aligned}&\min_{x}c^{\intercal}x\\ &\text{s.t.}Ax=b\\ &x\succeq 0\end{aligned}$$

In general with A rows a_i and columns A_j

$$\begin{array}{llllll}\min_x & c^{\intercal}x & & \max_p & p^{\intercal}b & \\ \text{s.t.} & a_ix\geq b_i & i\in M_1 & & p_i\geq 0 & i\in M_1\\ & a_ix\leq b_i & i\in M_2 & & p_i\leq 0 & i\in M_2\\ & a_ix=b_i & i\in M_3 & & p_i \text{ free} & i\in M_3\\ & x_j\geq 0 & j\in N_1 & & p^{\intercal}A_j\leq c_j & j\in N_1\\ & x_j\leq 0 & j\in N_2 & & p^{\intercal}A_j\geq c_j & j\in N_2\\ & x_j \text{ free} & j\in N_3 & & p^{\intercal}A_j=c_j & j\in N_3\end{array}$$

Why? Consider

$$\begin{aligned}&\min_{x}c^{\intercal}x\\ &\text{s.t.}Ax\preceq b\\ &x\succeq 0\end{aligned}$$

Then

$$\begin{aligned}&\min_{x}c^{\intercal}x\\ &\text{s.t.}0\preceq(b-Ax)\\ &x\succeq 0\end{aligned}$$

and

$$g(\lambda)=\min_{x\succeq 0}\left[c^{\intercal}x+\lambda^{\intercal}(b-Ax)\right]\leq c^{\intercal}x^*+\lambda^{\intercal}(b-Ax^*)\leq c^{\intercal}x^*$$

iff $\lambda\preceq 0$ since $(b-Ax^*)\succeq 0$. Then the rest of the proof goes through the same. What about if

$$\begin{aligned}&\min_{x}c^{\intercal}x\\ &\text{s.t.}Ax\preceq b\\ &x\preceq 0\end{aligned}$$

Then

$$\min_{x\preceq 0}\left(c^{\intercal}-\lambda^{\intercal}A\right)x=\begin{cases}-\infty & \text{if } \left(c^{\intercal}-\lambda^{\intercal}A\right)\not\preceq 0\\ 0 & \text{o/w}\end{cases}$$

i.e. if there exists a component of $(c^\intercal - \lambda^\intercal A)$ that's positive (because then we could crank that components to negative infinity). Therefore $(c^\intercal - \lambda^\intercal A) \preceq 0$ or

$$\max_\lambda \lambda^\intercal b$$
$$\text{s.t.}\,(c^\intercal - \lambda^\intercal A) \preceq 0$$

or

$$\max_\lambda \lambda^\intercal b$$
$$\text{s.t.}\, c^\intercal \preceq \lambda^\intercal A$$

2.4. **Supporting hyperplanes.** A hyperplane supports a set if it intersects the set and the entire set is on one side of the set.

Theorem. *Supporting hyperplane theorem: if S is a convex set then there exists a supporting hyperplane. I.e. for every $\bar{x} \in \partial S$ there exists $p \neq 0$ such that $p^\intercal (x - \bar{x}) \leq 0$ for all $x \in cl(S)$.*

Proof. Since $\bar{x} \in \partial S$ there exists a sequence $\{y_k\}$ not in $cl(S)$ such that $y_k \to \bar{x}$ (since ∂S is also the boundary of the complement of S). By the point separation theorem there exists a p_k (that we can normalize) such that

$$p_k^\intercal (y_k - x) > 0 \iff p_k^\intercal y_k > p_k^\intercal x$$

for each $x \in cl(S)$. Since $\{p_k\}$ are bounded there exists a convergent subsequence p_{k_j} with a limit p whose norm is equal to 1. Taking both limits simultaneously we get that $p^\intercal (\bar{x} - x) \geq 0$ or $p^\intercal (x - \bar{x}) \leq 0$. Basically you want to construct the hyperplane that goes through the point on the boundary but using the point set separation theorem takes using a limit to hone in on it (i.e. the plane should have wellbehaved properties under taking limit of all the planes for points not in). □

Corollary. *For a nonempty convex set S if $x \notin int(S)$ then there exists a separating hyperplane.*

Proof. If $x \notin cl(S)$ then just use point set separation. Otherwise just use the immediately previous theorem. □

Corollary. *For a nonempty S and $y \notin int(conv(S))$ there exists a separating hyperplane that separates S and y.*

Proof. By immediately prior since $conv(S)$ is convex. □

Corollary. *Let S_1, S_2 be two nonempty convex sets such that $S_1 \cap S_2 = \emptyset$. Then there exists a separating hyperplane, i.e. there exists p such that*

$$\inf\left\{p^\intercal x \middle| x \in S_1\right\} \geq \sup\left\{p^\intercal x \middle| x \in S_2\right\}$$

Proof. Let $S = S_1 \ominus S_2$. Note that S is convex and $0 \in S$ (since otherwise $S_1 \cap S_2$ would nonempty). By the first corollary there exists a separating hyperplane between S and 0, i.e. $p^\intercal (0 - x) \leq 0$ which is the same as $p^\intercal x \geq 0$ for all $x \in S$ which is the same

$$p^\intercal x_1 \geq p^\intercal x_2$$

for all $x_1 \in S_1$ and $x_2 \in S_2$. □

Corollary. *Let S_1, S_2 be two nonempty convex sets such that $S_1 \cap int(S_2) = \emptyset$ and $int(S_2) \neq \emptyset$. Then there exists a separating hyperplane.*

Proof. Interior of nonempty convex sets are convex so apply the previous result. □

Corollary. *Let S_1, S_2 be two nonempty closed convex sets such that S_1 is bounded and $S_1 \cap S_2 = \emptyset$. Then there exists a separating hyperplane that strongly separates.*

3. Inner Representation of convex sets

3.1. **Extreme points.** Consider the polyhedral set $S = \left\{x \middle| Ax = b, x \succeq 0\right\}$ where $A \in \mathbb{R}^{m \times n}$. Assume $\operatorname{rank}(A) = m$. If not, assuming $Ax = b$ is consistent, you can throw away linearly dependent rows in order to get a full row rank matrix. Rearrange the columns of A so that $A = [B, N]$ where $B \in \mathbb{R}^{m \times m}$ and full rank and $N \in \mathbb{R}^{m \times (n-m)}$ is the rest of the matrix. Then

$$\begin{aligned} Ax = Bx_B + Nx_N &= b \\ x_B &\succeq 0 \\ x_N &\succeq 0 \end{aligned}$$

Theorem. *x is an extreme point of S iff A can be decomposed into $[B, N]$ such that*

$$x = \begin{bmatrix} x_B \\ x_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

Proof. Why would this be true? A vertex (i.e. extreme point) of a polyhedron is a unique solution to a set of constraints. If it weren't unique then it would be free (a line) and therefore not a vertex. The question is how many constraints? For a polyhedron in $\mathbb{R}^n$ it has to be at least n constraints (a "point" with $n-1$ entries constrained and 1 free is a line in $\mathbb{R}^n$). And the constraints have to be linearly independent (otherwise you could get rid of redundancies and you'd only be satisfying $n' < n$ constraints). But could there be more? For example 3 lines intersecting? There's definitely a unique point satisfying that (if no two are colinear) but one of the lines is linearly dependent on the other two, i.e. redundant. More than n equations can't be linearly independent in $\mathbb{R}^n$. Therefore for a polyhedron in n variables, i.e. some number of constraints on $x \in \mathbb{R}^n$, we need at least n constraints to force a unique solution. If the rank of A is m then we can get a unique solution to m constraints by solving that system of m equations. Where do we get the rest from? We set the positivity constraints to be tight, i.e. $x_i = 0$. This partitioning of $x = [x_B, x_N] = [x_B, 0]$ effects exactly this. Such solutions x are called **basic feasible solutions**.

$\Leftarrow$Suppose that A can decomposed $A = [B, N]$ with $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$ and $x_B \succeq 0$. Then immediately $x \in S$. Suppose, towards a contradiction, that $x = \lambda x_1 + (1-\lambda) x_2$ for some $x_1, x_2 \in S$ for some $\lambda \in (0,1)$, i.e. x is not an extreme point. Then

$$x_1 = \begin{bmatrix} x_{1B} \\ x_{1N} \end{bmatrix}, x_2 = \begin{bmatrix} x_{2B} \\ x_{2N} \end{bmatrix}$$

and

$$\begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} x_{1B} \\ x_{1N} \end{bmatrix} + (1-\lambda) \begin{bmatrix} x_{2B} \\ x_{2N} \end{bmatrix}$$

Then $0 = \lambda x_{1N} + (1-\lambda) x_{2N}$ and $\lambda \in (0,1)$ forces $x_{1N} = x_{2N} = 0$. But then $\lambda x_{1B} + (1-\lambda) x_{2B} = B^{-1}b$ and $B^{-1}b$ is unique and hence $x_{1B} = x_{2B}$ and $x_1 = x_2$ and x is therefore an extreme point.

$\Rightarrow$Suppose x is an extreme point (vertex). Without loss of generality $x = (x_1, \dots, x_k, 0, \dots, 0)$ where $x_i \geq 0$ (we allow $k = n$). Firstly $a_1, \dots, a_k$ are linearly independent:

Proof. Towards a contradiction, suppose $\sum_{j=1}^k \lambda_j a_j = 0$ with λ_j not all equal to 0. Let

$$\lambda = (\lambda_1, \dots, \lambda_k, 0, \dots, 0)$$

and $\alpha > 0$ such that $x_1, x_2 \succeq 0$ and

$$x_1 = x + \alpha\lambda \text{ and } x_2 = x - \alpha\lambda$$

Note that

$$Ax_1 = Ax + \alpha A\lambda = Ax + \alpha \sum_{j=1}^k \lambda_i a_j = b$$

and similarly $Ax_2 = b$. Therefore $x_1, x_2 \in S$ and since $\alpha > 0$ and $\lambda \neq 0$ and $x_1 \neq x_2$ and $x = (1/2)\, x_1 + (1/2)\, x_2$ contradicting that x is an extreme point. $\square$

Thus $a_1, \ldots, a_k$ are linearly independent and since A has rank m, $m-k$ of the last $n-k$ columns may be chosen such that they, together with the first k columns, form a linearly independent set of m vectors; suppose $a_{k+1}, \ldots, a_m$ are these columns. Therefore

$$\begin{aligned} A &= [[a_1, \ldots, a_m], N] \\ &= [B, N] \end{aligned}$$

where B is full rank m. Furthermore for $x = (x_1, \ldots, x_k, 0)$

$$Ax = Bx + N0 = b$$

and therefore $(x, 0) = \left(B^{-1}b, 0\right)$ and since $x_j > 0$ it's the case that $B^{-1}b \succeq 0$. □

Corollary. *The number of extreme points of a polyhedron is less than or equal to*

$$\binom{n}{n-m} = \binom{n}{m}$$

because you can choose $n-m$ constraints to set to 0.

Corollary. *Let $S = \left\{x \middle| Ax = b, x \succeq 0\right\}$ be nonempty and $A \in \mathbb{R}^{m \times n}$ and rank$(A) = m$. Then S has at least one extreme point.*

Proof. Let $x \in S$, i.e. $Ax = b$, and without loss of generality, suppose that $x = (x_1, \ldots, x_k, 0, \ldots, 0)$ where $x_j > 0$. If $a_1, \ldots, a_k$ are linearly independent then $k \leq m$ and x is an extreme point (since x is a unique solution to $A(x_1, \ldots, x_k, 0) = b$). Otherwise $\sum_{j=1}^{k} \lambda_j a_j = 0$. Let

$$\alpha = \min_{1 \leq j \leq k} \left\{ \frac{x_j}{\lambda_j} \middle| \lambda_j > 0 \right\} = \frac{x_i}{\lambda_i}$$

Consider the point x' such that

$$x_j^{'} = \begin{cases} x_j - \alpha\lambda_j & j = 1, \ldots, k \\ 0 & j = k+1, \ldots, n \end{cases}$$

Note that $x_i^{'} = 0$ and

$$\sum_{j=1}^{n} a_j x_j^{'} = \sum_{j=1}^{n} a_j (x_j - \alpha\lambda_j) = b - 0 = b$$

Thus x' is feasible and has at most $k-1$ positive components. Repeat the process until the number of components corresponds to the number of linearly independent columns in A and you have an extreme point. □

Every basic feasible solution corresponds to an extreme point but an extreme point might be represented by several extreme points. The number of extreme points of a polytope is $\binom{n}{m}$ where $\text{rank}(A) = m$.

A **recession direction** is d such that $x + \lambda d \in P$ for all $\lambda \geq 0$. A recession direction is one such that

$$d \neq \lambda_1 d_1 + \lambda_2 d_2$$

for any distinct directions. The set of all recession directions is a cone. If $P = \left\{x \middle| Ax = b, x \succeq 0\right\}$ then $rec(P) = \left\{d \middle| Ad = 0, x \succeq 0\right\}$, since

$$A(x + \lambda d) = Ax + \lambda Ad) = b$$

for all λ necessitates that $Ad = 0$. Also d is a recession direction iff it's in the recession cone. To find extreme directions find the extreme points (i.e. basic feasible solutions of the system $Ad = 0$). An extreme ray is the entire ray while an extreme direction is just the direction.

Theorem. *Characterization of extreme directions. Let $S = \{x | Ax = b, x \succeq 0\}$ be nonempty and $A \in \mathbb{R}^{m\times n}$ and rank$(A) = m$. Then $\bar{d}$ is an extreme direction of S iff $A = [B, N]$ such that $B^{-1}a_j \preceq 0$ for some column of N and*

$$\bar{d} = \lambda \begin{pmatrix} -B^{-1}a_j \\ e_j \end{pmatrix}$$

with $\lambda > 0$.

Theorem. *Minkowski's theorem (representation theorem): every point in a polyhedron can be represented as the* ***convex*** *combination of extreme points plus* ***conic*** *combination of extreme directions, i.e.*

$$x = \sum_{j=1}^{k} \lambda_j x_j + \sum_{j=1}^{l} \mu_j d_j$$

where x_j are extreme points and d_j are extreme directions and $\sum_{j=1}^{k} \lambda_j = 1$ and $\mu_j, \lambda_j \geq 0$. This is the inner representation of the polygon.

Corollary. *If S is nonempty of the form $\{x | Ax = b, x \succeq 0\}$, i.e. a nonempty polyhedron. Then S has at least one extreme direction iff it's unbounded.*

Proof. If S has no extreme directions then by Minkowski's representation theorem (and Cauchy-Schwartz since $\lambda_j \geq 0$)

$$\|x\| = \left\| \sum_{j=1}^{k} \lambda_j x_j \right\| \leq \sum_{j=1}^{k} \lambda_j \|x_j\| = \sum_{j=1}^{k} \|x_j\|$$

for all $x \in S$. Therefore S is bounded. If S has an extreme direction then obviously it's unbounded. $\square$

Corollary. *The linear program $\mathcal{P}$*

$$\begin{aligned} &\min_x c^{\intercal} x \\ &s.t. Ax = b \\ &\quad x \succeq 0 \end{aligned}$$

with nonempty feasible region. Let $\{x_j\}$ be the set of extreme points of the feasible region and $\{d_j\}$ be the set of extreme directions.

(1) $\mathcal{P}$ has a finite optimal solution iff $c^{\intercal} d_j \geq 0$, i.e. the objective normal makes an acute angle with each extreme direction. Why does this make sense? For a minimization LP c is opposite of the direction in which the objective increases. If there exists $c^{\intercal} d_j < 0$ then $(-c)^{\intercal} d_j > 0$ and therefore going in the direction d_j decreases the objective arbitrarily.
(2) If no extreme directions ruin it then there exists an extreme point x_j that's optimal.

Proof. By representation theorem $Ax = b$ and $x \succeq 0$ iff

$$\begin{aligned} x &= \sum_{j=1}^{k} \lambda_j x_j + \sum_{j=1}^{l} \mu_j d_j \\ &\in \operatorname{conv}(x_1, \dots, x_k) \cup \operatorname{coni}(d_1, \dots, d_l) \end{aligned}$$

i.e. $\sum_{j=1}^{k} \lambda_j = 1$, $\lambda_j \geq 0$, $\mu_j \geq 0$ and the LP ca be re-expressed as

$$\begin{aligned}
&\min_{\boldsymbol{\lambda},\boldsymbol{\mu}} c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_j + \sum_{j=1}^{l} \mu_j d_j \right) \\
&\text{s.t.} \sum_{i=1}^{k} \lambda_i = 1 \\
&\quad\ \lambda_i \geq 0 \\
&\quad\ \mu_j \geq 0
\end{aligned}$$

So if $c^{\mathsf{T}} d_q < 0$ for some q then

$$c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_j + \sum_{j=1}^{l} \mu_j d_j \right) = c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_j + \sum_{\substack{j=1 \\ j \neq q}}^{l} \mu_j d_j \right) + \mu_q \left(c^{\mathsf{T}} d_q \right)$$

and so u_q could be chosen arbitrarily large in order to decrease the objective. So feasibility iff $c^{\mathsf{T}} d_j \geq 0$ for all $j = 1, \ldots, l$. Therefore

$$\min_{\boldsymbol{\lambda}} c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_j \right) \leq \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_j + \sum_{j=1}^{l} \mu_j d_j \right)$$

(again since $\mu_j c^{\mathsf{T}} d_j \geq 0$ for all j), i.e. choose $\mu_j = 0$ for all j. Thus

$$\min_{\boldsymbol{\lambda}} c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_j \right)$$

Clearly this is minimized by choosing $\lambda_q = 1$ for q such that $c^{\mathsf{T}} x_q = \min_{1 \leq j \leq k} c^{\mathsf{T}} x_j$. Why?

$$\begin{aligned}
\min_{\boldsymbol{\lambda}} c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_j \right) &\geq \min_{\boldsymbol{\lambda}} c^{\mathsf{T}} \left(\sum_{j=1}^{k} \lambda_j x_q \right) \\
&= \min_{\boldsymbol{\lambda}} c^{\mathsf{T}} x_q \left(\sum_{j=1}^{k} \lambda_j \right) \\
&= c^{\mathsf{T}} x_q \times 1
\end{aligned}$$

The smallest weighted average is found by putting all of the weight on the smallest object (since all the objects are positive). $\square$

3.2. **Exposed points.**

Definition. Let C be a nonempty closed convex set in $\mathbb{R}^n$. $x^* \in C$ is called an exposed solution if there exists a linear objective $f(x) = c^t x$ for which $x^* = \min_{x \in C} f(X)$.

Theorem. *Straszewicz's theorem. For any closed convex set C, the set of exposed solutions of C is a dense subset of the set of extreme points of C. Thus every extreme point of C is the limit point of some sequence of exposed points.*

Corollary. *Any closed bounded convex set C can be expressed as the closure of the convex hull of its exposed points.*

4. Convex functions

Let S be a nonempty convex subset of $\mathbb{R}^n$

Definition. A function $f : S \to \mathbb{R}$ is said to be *convex* if for all $x_1, x_2 \in S$ and $\lambda \in [0,1]$

$$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Definition. A function $f : S \to \mathbb{R}$ is said to be *strictly convex* if for all $x_1, x_2 \in S$, $x_1 \neq x_2$ and $\lambda \in (0,1)$

$$f(\lambda x_1 + (1-\lambda) x_2) < \lambda f(x_1) + (1-\lambda) f(x_2)$$

A strictly convex function basically has no linear pieces.

Facts:

(1) $\{f_i\}$ convex a conic $\alpha_j > 0$ combination $f(\mathbf{x}) = \sum_{j=1}^{k} \alpha_j f_j(\mathbf{x})$.
(2) g concave then $f : \{x | g(x) > 0\} \to \mathbb{R}$, i.e. $f(x) = 1/g(x)$ is convex.
(3) g be nondecreasing, convex, and h convex then $f(x) = g(h(x))$ is convex. g has to be nondecreasing!
(4) g be convex and $h(x) = Ax + b$ then $f(x) = g(Ax+b)$ is convex.

Theorem. *Multivariable f is convex iff f is convex on any line, i.e. $F_{\bar{x},d}(\lambda) = f(\bar{x} + \lambda d)$ is convex for all $\bar{x}, d \in \mathbb{R}^n$ as a function of λ.*

Theorem. *Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a convex function. Then the α-level-set of f is a convex for each $\alpha \in \mathbb{R}$, i.e. $S_\alpha = \{x \in S | f(x) \leq \alpha\} \subset \mathbb{R}^n$ is convex. The α-level-set is in the domain of the function.*

Proof. Let $x_1, x_2 \in S_\alpha$. Thus $x_1, x_2 \in S$ and $f(x_i) \leq \alpha$. Then

$$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha$$

□

Definition. Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$, then $\{(x, f(x)) | x \in S\} \subset \mathbb{R}^{n+1}$ is the *graph* of f.

Definition. Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$ and $S \neq \emptyset$. The *epigraph* of f, denoted $\mathrm{epi}(f)$, $\{(x,y) | y \geq f(x), x \in S\} \subset \mathbb{R}^{n+1}$. The *hypograph* of f, denoted $\mathrm{hypo}(f)$, $\{(x,y) | y \leq f(x), x \in S\} \subset \mathbb{R}^{n+1}$.

Theorem. *Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a convex function. Then f is convex iff $\mathrm{epi}(f)$ is convex.*

Theorem. *Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a convex function. Then f is continuous on the interior of S. But only the interior!*

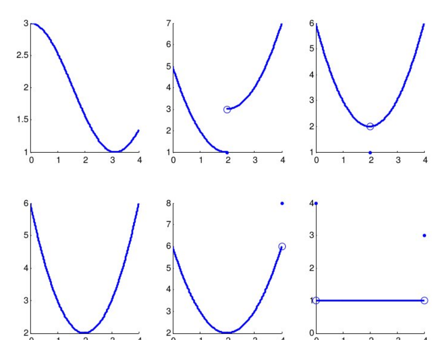

Figure 4.1. Convex functions are continuous on interiors of convex sets.

Definition. Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a convex function. Then ξ is a *subgradient* of f at $\bar{x}$ if

$$f(x) \geq f(\bar{x}) + \xi^t (x - \bar{x})$$

I.e. $f(x)$ is above the plane defined by $f(\bar{x}) + \xi^t(x - \bar{x})$. ξ is the **slope of the line or the gradient**.

Definition. The set of all subgradients $\partial f(\bar{x})$ of f at $\bar{x}$ is called the *subdifferential* of f at $\bar{x}$.

Theorem. *Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a convex function. Then f has a subgradient ξ at $\bar{x} \in int(S)$. In particular, the hyperplane*

$$\mathcal{H} = \left\{(x, y) \,\middle|\, y = f(\bar{x}) + \xi^t (x - \bar{x})\right\}$$

supports $epi(f)$ at $(\bar{x}, f(\bar{x}))$.

Proof. Note that epi (f) is convex and $(\bar{x}, f(\bar{x}))$ belongs to its boundary. Therefore by separation theorem there exists $(\xi_0, \mu) \in \mathbb{R}^n \times \mathbb{R}$. For all $(\mathbf{x}, y) \in$ epi (f).

$$\xi_0^t (x - \bar{x}) + \mu (y - f(\bar{x})) \leq 0$$

Note that $\mu \leq 0$ because otherwise take y large enough and the inequality would be violated. In fact $\mu < 0$:

Proof. Toward a contradiction suppose $\mu = 0$. Then $\xi_0^t (x - \bar{x}) \leq 0$ for all $x \in S$. Since $\bar{x} \in$ int (S) there exists $\lambda > 0$ such that $x = \bar{x} + \lambda \xi_0 \in S$ and hence $\lambda \xi_0^t \xi_0 \leq 0$. This implies that $\xi_0 = 0$ (otherwise how could $\lambda \xi_0^t \xi_0$ negative or equal to zero with λ strictly positive). But then $(\xi_0, \mu) = 0$ contradicting the separation theorem. Therefore $\mu < 0$. □

Dividing $\xi_0^t (x - \bar{x}) + \mu (y - f(\bar{x})) \leq 0$ by $|\mu|$

$$\xi^t (x - \bar{x}) + -1 (y - f(\bar{x})) = \xi^t (x - \bar{x}) - y + f(\bar{x}) \leq 0$$

for all $(x, y) \in$ epi (f). Then by letting $y \to f(x)$ we satisfy the theorem

$$\xi^t (x - \bar{x}) + f(\bar{x}) \leq f(x)$$

In particular, the hyperplane $H = \left\{(x, y) \,\middle|\, y = f(\bar{x}) + \xi^t (x - \bar{x})\right\}$ supports epi (f) at $(\bar{x}, f(\bar{x}))$. □

Corollary. *Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a strictly convex function. Then for $\bar{x} \in int(S)$, there exists a vector ξ such that*

$$f(x) > f(\bar{x}) + \xi^t (x - \bar{x})$$

Theorem. *Partial converse: let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$. Suppose for each $\bar{x} \in int(S)$ there exists ξ*

$$f(x) \geq f(\bar{x}) + \xi^t (x - \bar{x})$$

for all $x \in S$. Then f is convex on $int(S)$.

Definition. The *directional derivative* is

$$f'(\bar{x}, d) = \lim_{\lambda \to 0^+} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda}$$

Alternatively it's $d \cdot \nabla f(\bar{x})$

Theorem. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then f has all directional derivatives.*

Proof. Let $\lambda_2 > \lambda_1 > 0$. By convexity of f we have

$$\begin{aligned} f(\bar{x} + \lambda d) &= f\left(\frac{\lambda_1}{\lambda_2}(\bar{x} + \lambda_2 d) + \left(1 - \frac{\lambda_1}{\lambda_2}\right)\bar{x}\right) \\ &\leq \frac{\lambda_1}{\lambda_2} f(\bar{x} + \lambda_2 d) + \left(1 - \frac{\lambda_1}{\lambda_2}\right) f(\bar{x}) \end{aligned}$$

which implies

$$\frac{f(\bar{x}+\lambda_1 d)-f(\bar{x})}{\lambda_1} \leq \frac{f(\bar{x}+\lambda_2 d)-f(\bar{x})}{\lambda_2}$$

Thus the difference quotient is monotonically decreasing. Then since the function is convex it has a subgradient at $\bar{x}$ and so bounded below. Therefore the limit converges. $\square$

Definition. A function is called differentiable if there exists v such that

$$f(x) = f(\bar{x}) + v^t (x-\bar{x}) + \alpha(\bar{x}, x-\bar{x}) \|x-\bar{x}\|$$

The function α is the Lagrange form of the first order remainder term in the Taylor series approximation of f, i.e.

$$\alpha(\bar{x}, x-\bar{x}) \|x-\bar{x}\| = \frac{f''(\bar{x})}{3!} (x-\bar{x})^2$$

v is called the gradient and duh is $\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \ldots, \frac{\partial f(\bar{x})}{\partial x_n}\right)$.

Lemma. *Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a convex function and differentiable at $\bar{x}$. Then the subdifferential at $\bar{x}$ is the singleton $\{\nabla f(\bar{x})\}$.*

Proof. Since f is a convex set the subdifferential at $\bar{x}$ is not empty. Let ξ be a subgradient of f at $\bar{x}$. Again by the same theorem (existence of subgradients)

$$f(\bar{x}+\lambda d) \geq f(\bar{x}) + \xi^t (\lambda d)$$

By differentiability at $\bar{x}$

$$f(\bar{x}+\lambda d) = f(\bar{x}) + (\nabla f(\bar{x}))^t (\lambda d) + \alpha(\bar{x}, \lambda d) \|\lambda d\|$$

Subtracting the equation from the inequality we get that

$$0 \geq \xi^t (\lambda d) - (\nabla f(\bar{x}))^t (\lambda d) - \alpha(\bar{x}, \lambda d) \|\lambda d\|$$

Dividing by λ and letting $\lambda \to 0^+$ we get that $(\xi - \nabla f(\bar{x}))^t d \leq 0$. Since this is true for all d, choosing $d = \xi - \nabla f(\bar{x})$ proves that $\xi - \nabla f(\bar{x}) = 0$ (how else would the norm squared of it be nonpositive). $\square$

In light of the lemma, and supporting hyperplane, and the partial converse we have another characterization of convex functions:

Theorem. *Let S be a nonempty open (or $\bar{x} \in int(S)$) convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ differentiable. Then f is convex iff $\forall \bar{x} \in S$*

$$f(x) \geq f(\bar{x}) + (\nabla f(\bar{x}))^t (x-\bar{x})$$

Theorem. *Let S be a nonempty open (or $\bar{x} \in int(S)$) convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ differentiable. Then f is convex iff $\forall x_1, x_2 \in S$*

$$(\nabla f(x_2) - \nabla f(x_1))^t (x_2 - x_1) \geq 0$$

Proof. By characterization of convexity we have that

$$f(x_1) \geq f(x_2) + (\nabla f(x_2))^t (x_1 - x_2)$$
$$f(x_2) \geq f(x_1) + (\nabla f(x_1))^t (x_2 - x_1)$$

Adding the two inequalities gets the result. To prove the converse use the mean value theorem:

$$f(x_2) - f(x_1) = (\nabla f(x))^t (x_2 - x_1)$$

where $x = \lambda x_1 + (1-\lambda) x_2$ for $\lambda \in (0,1)$. By assumption $(\nabla f(x) - \nabla f(x_1))^t (x - x_1) \geq 0$ which is equivalent to

$$(1-\lambda)(\nabla f(x) - \nabla f(x_1))^t (x_2 - x_1) \geq 0$$

This implies that

$$(\nabla f(x))^t (x_2 - x_1) \geq (\nabla f(x_1))^t (x_2 - x_1)$$

But by mean value theorem result we get that

$$f(x_2) - f(x_1) \geq (\nabla f(x_1))^t (x_2 - x_1)$$

and so by previous characterization we get that f is convex. $\square$

This is a first order condition on convexity. How about second order conditions?

Definition. A function f is twice differentiable if

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^t H(\bar{x})(x - \bar{x}) + \alpha(\bar{x}, x - \bar{x}) \|x - \bar{x}\|^2$$

and $\lim_{x \to \bar{x}} \alpha(\bar{x}, x - \bar{x}) = 0$.

Theorem. *Let S be a nonempty open (or $\bar{x} \in int(S)$) convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ twice differentiable. Then f is convex iff $\nabla^2 f(\bar{x})$ is psd.*

For strict convexity you need something stronger.

Theorem. *Let S be a nonempty open (or $\bar{x} \in int(S)$) convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ twice differentiable. Then*

(1) $\nabla^2 f(\bar{x})$ is pd then f is strictly convex.
(2) f is strictly convex then $\nabla^2 f(\bar{x})$ is psd.
(3) f is strictly convex and quadratic then $\nabla^2 f(\bar{x})$ is pd.

5. Optimality conditions for Convex programs

Definition. Consider

$$\min_x f(x)$$
$$\text{s.t.} x \in S$$

(1) $\bar{x} \in S$ is a **feasible solution** to the problem.
(2) $\bar{x} \in S$ such that $f(\bar{x}) \leq f(x)$ for all $x \in S$ is a **globally optimal solution**.
(3) $\bar{x} \in S$ such that $f(\bar{x}) \leq f(x)$ for all $x \in S \cap \mathcal{N}_\epsilon(\bar{x})$, i.e. in some neighborhood of $\bar{x}$ is a **locally optimal solution**.
(4) $\bar{x} \in S$ such that $f(\bar{x}) < f(x)$ for all $x \in S \cap \mathcal{N}_\epsilon(\bar{x})$, i.e. in some neighborhood of $\bar{x}$ is a **strict locally optimal solution**.
(5) $\bar{x} \in S$ such that $f(\bar{x}) < f(x)$ for all $x \in S \cap \mathcal{N}_\epsilon(\bar{x})$, i.e. in some neighborhood of $\bar{x}$ and $\bar{x}$ is the only such point is an **isolated locally optimal solution**.

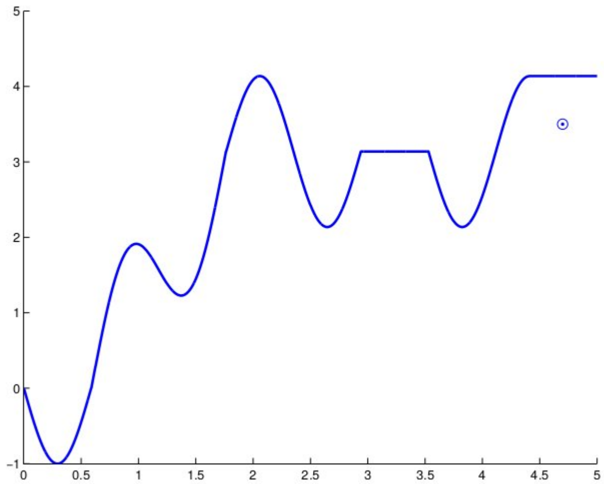

FIGURE 5.1. Characterizing solutions

Theorem. *Let S be a nonempty convex set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ convex and $\bar{x}$ be a locally optimal solution (optimal, but not uniquely, within an ϵ-ball) to $\min_{x \in S} f(x)$. Then*

(1) $\bar{x}$ is globally optimal.
(2) If $\bar{x}$ is strictly locally optimal ($f(\bar{x}) < f(x)$ in the ϵ-ball). or f is strictly convex then
  (a) $\bar{x}$ is uniquely globally optimal.
  (b) $\bar{x}$ is strongly (isolated) locally optimal.

Proof. Towards a contradiction suppose $\bar{x}$ is not globally optimal. Note $\bar{x}$ locally optimal means $f(\bar{x}) \le f(x)$ for $x \in S \cap \mathcal{N}_\epsilon(\bar{x})$ with $\epsilon > 0$. Then not being global means there exists $\hat{x}$ such that $f(\hat{x}) < f(\bar{x})$. By convexity of f

$$f(\lambda \hat{x} + (1-\lambda)\bar{x}) \le \lambda f(\hat{x}) + (1-\lambda) f(\bar{x})$$

Then using $f(\hat{x}) < f(\bar{x})$

$$f(\lambda \hat{x} + (1-\lambda)\bar{x}) \le \lambda f(\hat{x}) + (1-\lambda) f(\bar{x}) < \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) = f(\bar{x})$$

Taking $\lambda < \epsilon$ we get a contradiction. That handles part 1.

Part 2(a): Suppose $\bar{x}$ is a strict local optimal. Then by part 1 it's a global optimum. Suppose there exists $\hat{x}$ such that $f(\bar{x}) = f(\hat{x})$. Then

$$f(x_\lambda) = f(\lambda \hat{x} + (1-\lambda)\bar{x}) \le \lambda f(\hat{x}) + (1-\lambda) f(\bar{x}) = \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) = f(\bar{x})$$

and by taking $\lambda < \epsilon$ we get a contradiction to strict local optimality. Hence $\bar{x}$ is the unique global minimum.

Part 2(b): Since it's a unique global minimum it must be isolated since any other local minimum $\bar{x}$ in $S \cap \mathcal{N}_\epsilon(\bar{x})$ would be a global minimum.

Part 2'(a/b): Suppose $\bar{x}$ is a local optimum and f is strictly convex. Since strictly convex implies convexity $\bar{x}$ is still a global minimum. Towards a contradiction suppose $\bar{x}$ is not unique, i.e. there exists $\hat{x}$ such that $f(\hat{x}) = f(\bar{x})$. By strict convexity

$$f\left(\frac{\bar{x}+\hat{x}}{2}\right) < \frac{1}{2} f(\bar{x}) + \frac{1}{2} f(\hat{x}) = f(\bar{x})$$

By convexity of S we have $\frac{\bar{x}+\hat{x}}{2} \in S$ and therefore a contradiction of global optimality. Further similarly $\bar{x}$ is also isolated. $\square$

Theorem. *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex function and $S\neq\emptyset$ a convex subset of $\mathbb{R}^n$ and $\bar{x}\in S$. Then for*

$$\min_{x\in S} f(x)$$

$\bar{x}$ is globally optimal iff f has a supporting subgradient ξ such that $\xi^t(x-\bar{x})\geq 0$ for all $x\in S$.

Proof. $\Leftarrow$ Suppose there exists ξ such that $\xi^t(x-\bar{x})\geq 0$ for all $x\in S$. By convexity of f

$$f(x)\geq f(\bar{x})+\xi^t(x-\bar{x})\geq f(\bar{x})$$

since $\xi^t(x-\bar{x})$. Hence $\bar{x}$ is optimal.

$\Rightarrow$ Suppose $\bar{x}$ is optimal. Let

$$\begin{aligned}\Lambda_1&=\left\{(x-\bar{x},y)\,\middle|\,x\in\mathbb{R}^n,y>f(x)-f(\bar{x})\right\}\\ \Lambda_2&=\left\{(x-\bar{x},y)\,\middle|\,x\in S,y\leq 0\right\}\end{aligned}$$

Note that Λ_1 is the "shifted" epigraph of f, shifted such that $(\bar{x},f(\bar{x}))=(0,0)$. Both Λ_1 and Λ_2 are convex. Also $\Lambda_1\cap\Lambda_2=\emptyset$ since otherwise there exists (x,y) such that $x\in S$ and $0>y>f(x)-f(\bar{x})$ contradicting $\bar{x}$ is optimal (since $0>f(x)-f(\bar{x})\iff f(\bar{x})>f(x)$). By separating hyperplane theorem there exists a hyperplane that separates: there exists nonzero (ξ_0,μ) and $\alpha\neq 0$ such that

$$\begin{aligned}&\xi_0^t(x-\bar{x})+\mu y\leq\alpha,\ \forall x\in\mathbb{R}^n,\ y>f(x)-f(\bar{x})\\ &\xi_0^t(x-\bar{x})+\mu y\geq\alpha,\ \forall x\in S,\ y\leq 0\end{aligned}$$

Then letting $x=\bar{x}$ and $y=0$ in the second inequality you get that $\alpha\leq 0$. Now letting $x=\bar{x}$ and $y=\epsilon>0$ in the first inequality and you get $\mu\epsilon\leq\alpha$. Since this is true for every $\epsilon>0$ it's the case that $\alpha\geq 0$ and therefore$\mu\leq 0$. Therefore $\mu\leq 0$ and $\alpha=0$. So summarizing

$$\begin{aligned}&\xi_0^t(x-\bar{x})+\mu y\leq 0,\ \forall x\in\mathbb{R}^n,\ y>f(x)-f(\bar{x})\\ &\xi_0^t(x-\bar{x})+\mu y\geq 0,\ \forall x\in S,\ y\leq 0\end{aligned}$$

If μ were 0 then $\xi_0^t(x-\bar{x})\leq 0$ for all $x\in\mathbb{R}^n$ and then letting $x=\bar{x}+\xi_0$ shows that $\xi_0=0$ which isn't possible. So $\mu<0$. Dividing by $|\mu|$ everywhere we get that

$$\begin{aligned}&\xi^t(x-\bar{x})-y\leq 0,\ \forall x\in\mathbb{R}^n,\ y>f(x)-f(\bar{x})\\ &\xi^t(x-\bar{x})-y\geq 0,\ \forall x\in S,\ y\leq 0\end{aligned}$$

Letting $y=0$ in the second inequality we get that $\xi^t(x-\bar{x})\geq 0$ for all $x\in S$. From the first inequality we conclude that since $\xi^t(x-\bar{x})-y\leq 0$ for all (x,y) in the "shifted" strict epigraph of f it must therefore also hold in the closure, i.e. where $y=f(x)-f(\bar{x})$, which gives us

$$f(x)\geq f(\bar{x})+\xi^t(x-\bar{x})$$

□

Corollary. *If S is open then $\bar{x}$ is global optimal iff $0\in\partial f(\bar{x})$*

Proof. $\bar{x}$ is optimal iff there exists ξ such that $\xi^t(x-\bar{x})\geq 0$. Since S is open take λ such that $x=\bar{x}-\lambda\xi\in S$ and then

$$-\lambda\|\xi\|^2\geq 0$$

□

Corollary. *If f differentiable (and convex) then*

(1) $\bar{x}$ is globally optimal iff $\nabla f(\bar{x})^t(x-\bar{x})\geq 0$
(2) If S is open then $\bar{x}$ is globally optimal iff $\nabla f(\bar{x})=0$

Proof. Obvious since $\partial f(\bar{x})=\{\nabla f(\bar{x})\}$. □

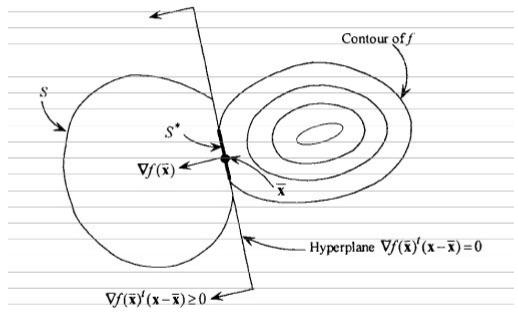

FIGURE 5.2. Gradient angle

Consider figure 5.2. Suppose the problem is to minimize f subject to $x \in S$ and f is differentiable and convex but S is arbitrary. Suppose at some $\bar{x}$ the directional derivative $\nabla f(\bar{x})(x-\bar{x}) \geq 0$ for all $x \in S$. Then going in any direction in S would potentially increase the objective, regardless of what S is like. Why? By convexity[2]and differentiability of f any solution $\hat{x}$ (anywhere in $\mathbb{R}^n$, which is a convex set) that improves on $\bar{x}$

$$f(\bar{x}) > f(\hat{x}) \geq f(\bar{x}) + \nabla f(\bar{x})^t (\hat{x}-\bar{x})$$

which implies $\nabla f(\bar{x})^t(\hat{x}-\bar{x}) < f(\bar{x}) - f(\bar{x}) = 0$ but $\nabla f(\bar{x})^t (x-\bar{x}) \geq 0$ for all $x \in S$. Hence the hyperplane $\nabla f(\bar{x})^t(\hat{x}-\bar{x}) = 0$ separates S (arbitrary S) from solutions that improve cost. In the non-differentiable case the supporting hyperplane ξ plays the same role as $\nabla f(\bar{x})$. Conversely suppose f is differentiable but arbitrary otherwise and S is convex. Then $\bar{x}$ is a global minimum then again $\nabla f(\bar{x})(x-\bar{x}) \geq 0$ since otherwise if there exists $x \in \operatorname{int}(S)$ such that $\nabla f(\bar{x})^t(x-\bar{x}) < 0$ you could do go in the direction $d = x - \bar{x}$, improve cost, and still satisfy constraints. The general take away is: if f is differentiable and otherwise f and S are arbitrary and $\bar{x}$ is a local minimum then for any feasible direction d such that $x = \bar{x} + \lambda d$ it must be the case that

$$\nabla f(\bar{x})^t d \geq 0$$

for some $0 < \lambda \leq \delta$ i.e. going in that direction for a small enough step does not improve the objective.

Back to your original program:

What characterizes the set of optimal solutions to $\min \left\{ f(x) \middle| x \in S \right\}$ when f is convex and differentiable and so is S?

Theorem. *If f is convex and differentiable and S is convex. Suppose there exists an optimal solution $\bar{x}$. Then the set of optimal solution S^**

$$S^* = \left\{ x \in S \middle| \nabla f(\bar{x})^t (x-\bar{x}) \leq 0, \nabla f(x) = \nabla f(\bar{x}) \right\}$$

Proof. Denote the candidate set of optimal solutions by $\bar{S}$ and note that $\bar{x} \in \bar{S}$. Consider $\hat{x} \in S^*$. By convexity of f (and convexity of S) and definition of S^*, it's the case that $\hat{x} \in S$

$$f(\bar{x}) \geq f(\hat{x}) + \nabla f(\hat{x})^t (\bar{x}-\hat{x}) = f(\hat{x}) + \nabla f(\bar{x})^t(\bar{x}-\hat{x}) = f(\hat{x}) + \left(-\nabla f(\bar{x})^t(\hat{x}-\bar{x})\right) \geq f(\hat{x})$$

Hence $\hat{x} \in \bar{S}$ and so $S^* \subset \bar{S}$. Conversely, suppose $\hat{x} \in \bar{S}$ then $f(\hat{x}) = f(\bar{x})$ and so

$$f(\bar{x}) = f(\hat{x}) \geq f(\bar{x}) + \nabla f(\bar{x})^t (\hat{x}-\bar{x})$$

and hence $\nabla f(\bar{x})^t(\hat{x}-\bar{x}) \leq 0$ but by corollary above $\nabla f(\bar{x})^t(\hat{x}-\bar{x}) \geq 0$ and hence $\nabla f(\bar{x})^t(\hat{x}-\bar{x}) = 0$. Interchanging $\hat{x}$ and $\bar{x}$ we get that $\nabla f(\bar{x})^t(\bar{x}-\hat{x}) = 0$ and subtracting we get

$$(\nabla f(\bar{x}) - \nabla f(\hat{x}))^t (\bar{x}-\hat{x}) = 0$$

[2]At every $\bar{x}$ we have that $f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^t (x-\bar{x})$.

But

$$\begin{aligned}(\nabla f(\bar{x}) - \nabla f(\hat{x})) &= \nabla f(\hat{x} + \lambda(\bar{x} - \hat{x}))\Big|_{\lambda=0}^{\lambda=1} \\ &= \int_0^1 \nabla^2 f(\hat{x} + \lambda(\bar{x} - \hat{x}))(\bar{x} - \hat{x})\, d\lambda =: G(\bar{x} - \hat{x})\end{aligned}$$

Note that G is psd since $\nabla^2 f$ is psd (since f is convex). Then

$$0 = (\bar{x} - \hat{x})^t (\nabla f(\bar{x}) - \nabla f(\hat{x})) = (\bar{x} - \hat{x})^t G(\bar{x} - \hat{x})$$

and by psd-ness $G(\bar{x} - \hat{x})$ must be 0 and hence $\nabla f(\bar{x}) - \nabla f(\hat{x}) = 0$ (and $\nabla f(\bar{x})^t (\hat{x} - \bar{x}) \leq 0$) and hence $\hat{x} \in S^*$ and so $\bar{S} \subset S^*$. ☐

Corollary. *The set of alternative solutions S^* for convex, twice differentiable, and convex S can be characterized as*

$$S^* = \left\{ x \in S \middle| \nabla f(\bar{x})^t (x - \bar{x}) = 0, \nabla f(x) = \nabla f(\bar{x}) \right\}$$

Proof. Follows from previous theorem and for optimal $\bar{x}$, and all other $x \in S$, $\nabla f(\bar{x})(x - \bar{x}) \geq 0$. ☐

Corollary. *Suppose $f = c^t x + \frac{1}{2} x^t H x$ and S is polyhedral. Then S^* is the polyhedral set given by*

$$S^* = \left\{ x \in S \middle| c^t (x - \bar{x}) = 0, H(x - \bar{x}) = 0 \right\}$$

Proof. $\nabla f(x) = c + Hx$ ☐

For maxima of convex functions similar things apply (but for different reasons).

Theorem. *If f is convex and S a nonempty convex set, if $\bar{x}$ is a local optimal then for all ξ it's the case that $\xi^t (x - \bar{x}) \leq 0$ for each $x \in S$.*

Proof. Suppose $\bar{x} \in S$ is a local optimum. Then for all $x \in S \cap \mathcal{N}_\epsilon(\bar{x})$ it's the case that $f(x) \leq f(\bar{x})$. Let $\lambda < \epsilon$ and then

$$f(\bar{x} + \lambda(x - \bar{x})) \leq f(\bar{x})$$

Then by convexity of f for all ξ

$$f(\bar{x} + \lambda(x - \bar{x})) \geq f(\bar{x}) + \lambda \xi^t (x - \bar{x})$$

implying $\lambda \xi^t (x - \bar{x}) \leq 0$. Divinding by λ we have it. ☐

Corollary. *If f is differentiable them $\nabla f(\bar{x})(x - \bar{x}) \leq 0$.*

The result is necessary but not sufficient for gradients to be as such.

Theorem. *If f is a convex function and S a nonempty polyhedral set then the solution to a maximization problem is at the boudary (i.e. extreme point).*

Proof. Note f is continuous because it's convex and S is compact (bounded over $\mathbb{R}^n$) and hence f achieves its maximum at $x' \in S$. By the representation theorem S is defined only as a convex combination of extreme points hence

$$x^{'} = \sum_{j=1}^{k} \lambda_j x_j, \sum_{j=1}^{k} \lambda_j = 1$$

and $\lambda_j \geq 0$ and x_j are extreme points. By convexity of f

$$f\left(\sum_{j=1}^{k} \lambda_j x_j \right) \leq \sum_{j=1}^{k} \lambda_j f(x_j)$$

But $f\left(x^{'}\right) \geq f(x_j)$ for all j and thus $f\left(x^{'}\right) = f(x_j)$. Therefore the solutions are all at the bounday points. ☐

Definition. Consider $\min_{x\in\mathbb{R}^n} f(x)$. Then d is an *improving* direction at $\bar{x}$ if $f(\bar{x}+\lambda d) < f(\bar{x})$ for $0<\lambda<\epsilon$ for some ϵ.

Theorem. *Let d be such that $\nabla f(\bar{x})^t d < 0$. Then d is an improving direction.*

Proof. Why? This says that going in the direction of is strictly acute with $-\nabla f(\bar{x})$, i.e. sort of aligned with the direction of steepest descent. By differentiability of f at $\bar{x}$

$$f(\bar{x}+\lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^t d + \lambda \|d\| \alpha(\bar{x}, \lambda d)$$

Re-arranging terms we get

$$\frac{f(\bar{x}+\lambda d) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^t d + \|d\| \alpha(\bar{x}, \lambda d)$$

Since $\nabla f(\bar{x})^t d < 0$ and $\alpha \to 0$ as $\lambda \to 0$ there exists ϵ such that $\nabla f(\bar{x})^t d + \|d\| \alpha(\bar{x}, \lambda d) < 0$ (we can make $\|d\| \alpha(\bar{x}, \lambda d)$ smaller than $\nabla f(\bar{x})^t d$) for all $0<\lambda<\epsilon$ and therefore

$$\frac{f(\bar{x}+\lambda d) - f(\bar{x})}{\lambda} < 0$$

and since $\lambda > 0$ we have that $f(\bar{x}+\lambda d) - f(\bar{x}) < 0$. □

Corollary. *If $\bar{x}$ is a local minimum then $\nabla f(\bar{x}) = 0$.*

Proof. Towards a contradiction suppose $\nabla f(\bar{x}) \neq 0$ and set $d = -\nabla f(\bar{x})$. Then $\nabla f(\bar{x})^t d = -\|\nabla f(\bar{x})\|^2 < 0$, hence satisfying the previous result for this direction, hence d is an improving direction contradicting that $\bar{x}$ is a local minimum. □

The converse is not true!

Second order necessary conditions

Theorem. *Let f be twice differentiable at $\bar{x}$. If $\bar{x}$ is a local minimum of f then $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is psd.*

Proof. First part follows from above. For the second part consider and arbitrary direction

$$f(\bar{x}+\lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^t d + \frac{1}{2}\lambda^2 d^t H(\bar{x}) d + \lambda^2 \|d\|^2 \alpha(\bar{x}, \lambda d)$$

Again $\nabla f(\bar{x}) = 0$ and so

$$\frac{f(\bar{x}+\lambda d) - f(\bar{x})}{\lambda^2} = +\frac{1}{2} d^t H(\bar{x}) d + \|d\|^2 \alpha(\bar{x}, \lambda d)$$

Since $\bar{x}$ is local minimum $f(\bar{x}+\lambda d) - f(\bar{x}) \geq 0$ for sufficiently small λ. Thus

$$\frac{1}{2} d^t H(\bar{x}) d + \|d\|^2 \alpha(\bar{x}, \lambda d) \geq 0$$

Taking the limit as $\lambda \to 0$ we get that $d^t H(\bar{x}) d \geq 0$ for all directions d. □

The converse is not true! But there is a partial converse

Theorem. *Let f be twice differentiable at $\bar{x}$. If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ pd then $\bar{x}$ is a strict local minimum.*

Proof. Since f is twice differentiable at $\bar{x}$

$$f\left(\bar{x}+\lambda d\right)=f\left(\bar{x}\right)+\lambda\nabla f\left(\bar{x}\right)^{t}d+\frac{1}{2}\lambda^{2}d^{t}H\left(\bar{x}\right)d+\lambda^{2}\left\|d\right\|^{2}\alpha\left(\bar{x},\lambda d\right)$$

Towards a contradiction suppose $\bar{x}$ is not a strict local minimum. We want find a direction along which $d^{t}Hd\leq 0$. Since $\bar{x}$ is not a strict local minimum there exists a sequence x_k such that $f\left(x_k\right)\leq f\left(\bar{x}\right)$. Denote

$$d_k=\frac{x_k-\bar{x}}{\left\|x_k-\bar{x}\right\|}$$

and noting that $\nabla f\left(\bar{x}\right)=0$ we have that

$$\frac{1}{2}d_k^{t}H\left(\bar{x}\right)d+\alpha\left(\bar{x},x_k-\bar{x}\right)\leq 0$$

But d_k is bounded (norm 1) and therefore there exists a subsequence that converges to a direction d such that $d^{t}Hd\leq 0$. □

In certain instances we can say stronger things:

Theorem. *Let f be* ***convex*** *and differentiable at $\bar{x}$. Then $\bar{x}$ is a global minimum iff $\nabla f\left(\bar{x}\right)=0$*

Proof. By corollary to two theorems ago if global minimum then $\nabla f\left(\bar{x}\right)=0$. For the converse suppose that $\nabla f\left(\bar{x}\right)=0$ so that $\nabla f\left(\bar{x}\right)^{t}\left(x-\bar{x}\right)=0$. By convexity

$$f\left(x\right)\geq f\left(\bar{x}\right)+\nabla f\left(\bar{x}\right)^{t}\left(x-\bar{x}\right)=f\left(\bar{x}\right)$$

□

7. Optimality conditions for constrained problems

Definition. Let S be a nonempty set in $\mathbb{R}^n$ and $\bar{x}\in\mathrm{cl}\left(S\right)$. The ***cone of feasible directions*** of S at $\bar{x}$, denoted by D, is given by

$$D=\left\{d\middle|\exists_{\delta}\forall_{0<\lambda<\delta}\,d\neq 0,x+\lambda d\in S\right\}$$

Definition. Given a function f and a minimization problem the ***cone of improving directions*** at $\bar{x}$ is

$$F=\left\{d\middle|f\left(\bar{x}+\lambda\right)<f\left(\bar{x}\right)\text{ for all }\lambda\in\left(0,\delta\right)\text{ for some }\delta\right\}$$

Definition. For differentiable f at $\bar{x}$ a subset of the improving directions (by theorem that says that $\nabla f\left(\bar{x}\right)^{t}d<0$ is an improving direction; d makes a strictly acute angle with $-\nabla f\left(\bar{x}\right)$ the direction of steepest descent) is

$$F_0=\left\{d\middle|\nabla f\left(\bar{x}\right)^{t}d<0\right\}$$

Theorem. *For* $\min\left\{f\left(x\right)\middle|x\in S\right\}$ *with f differentiable at $\bar{x}\in S$. If $\bar{x}$ is a local optimal solution then $F_0\cap D=\emptyset$.*

Proof. Towards a contradiction suppose $\bar{x}$ is locally optimal and there exists $d\in F_0\cap D$. Then by improving direction theorem ($\nabla f\left(\bar{x}\right)^{t}d<0\Rightarrow\exists_{\delta>0}\forall_{\lambda\in(0,\delta)}f\left(\bar{x}+\lambda d\right)<f\left(\bar{x}\right)$) there exists δ_1 such that for all $\lambda\in\left(0,\delta_1\right)$

$$f\left(\bar{x}+\lambda d\right)<f\left(\bar{x}\right)$$

Then by definition of D it's the case that there exists δ_2 (possibly different from δ_1) such that

$$\bar{x}+\lambda d\in S$$

Taking $\delta=\min\left\{\delta_1,\delta_2\right\}$ and we have that for $\lambda\in\left(0,\delta\right)$ it's the case that $\bar{x}+\lambda d\in S$ and

$$f\left(\bar{x}+\lambda d\right)<f\left(\bar{x}\right)$$

contradicting local optimality, a contradiction. □

There's a partial converse in the case of convex f

Theorem. *Suppose $F_0 \cap D = \emptyset$ and f is convex at $\bar{x}$ and there exists a neighborhood $\mathcal{N}_\epsilon(\bar{x})$ such that $(x - \bar{x}) \in D$ for any $x \in S \cap \mathcal{N}_\epsilon(\bar{x})$ (i.e. feasible direction in the neighborhood). Then $\bar{x}$ is a local minimum.*

Proof. Towards a contradiction suppose $F_0 \cap D = \emptyset$ and $x - \bar{x} \in D$ for all $x \in \mathcal{N}_\epsilon(\bar{x}) \cap S$ and there exists $\hat{x}$ such that $f(\hat{x}) < f(\bar{x})$ for some $\hat{x} \in \mathcal{N}_\epsilon(\bar{x}) \cap S$, i.e. $\bar{x}$ is not a local minimum. By assumption on $\mathcal{N}_\epsilon(\bar{x}) \cap S$ it's the case that $d = \hat{x} - \bar{x} \in D$. Furthermore by convexity $\nabla f(\bar{x})^t d < 0$ (Suppose $\nabla f(\bar{x})^t d \geq 0$ we would have $f(\hat{x}) = f(\bar{x} + d) \geq f(\bar{x}) + \nabla f(\bar{x})^t d \geq f(\bar{x})$). Therefore $d \in D$ and $d \in F_0$ which is a contradiction. $\square$

The cone F_0 is an algebraic description of the set of improving directions but $F_0 \subset F$. There's another inclusion: if $d \in F$ then $\nabla f(\bar{x})^t d \leq 0$ (since $\nabla f(\bar{x})^t d > 0$ would make d an ascent direction). Therefore

$$F_0 \subset F \subset F_0^{'} := \left\{ d \neq 0 \middle| \nabla f(\bar{x})^t d \leq 0 \right\}$$

The unknown directions are those for which $\nabla f(\bar{x})^t d = 0$. But in instances of convexity (concavity) we have tightness

Theorem. *If f is convex(concave) at $\bar{x}$ then $F = F_0$ ($F = F_0^{'}$).*

Proof. If f is convex at $\bar{x}$ then if $\nabla f(\bar{x})^t d \geq 0$ it's the case that $f(\bar{x} + \lambda d) \geq f(\bar{x}) + \nabla f(\bar{x})^t (\lambda d) \geq f(\bar{x})$ for all λ. Recall that

$$F = \left\{ d \middle| f(\bar{x} + \lambda) < f(\bar{x}) \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta \right\}$$

and hence negating the above we get that $d \in F$ implies $d \in F_0$, i.e. $F \subset F_0$ and so $F = F_0$. Now if f is strictly concave then we know that whenever $d \in F_0^{'}$ we have $f(\bar{x} + \lambda d) < f(\bar{x})$ and therefore $F_0^{'} \subset F$ and hence $F = F_0^{'}$. $\square$

Theorem. *From the improving directions theorem if $\nabla f(\bar{x})^t d < 0$ then d is an improving direction.*

Specify the feasible region using functions g_i now

$$S = \left\{ x \in X \middle| g_i(x) \leq 0 \right\}$$

We'd like to characterize D algebraically so that we can establish algebraic constraints on the objective and constraints for optimality.

Lemma. *Let*

$$S = \left\{ x \in X \middle| g_i(x) \leq 0, i \in M = \{1, \dots, m\} \right\}$$

Given a feasible $\bar{x}$ let $I \subset M$ be set of i such that $g_i(\bar{x}) = 0$ (the set of binding constraints) and assume g_I are differentiable at $\bar{x}$ and $M - I$ are continuous at $\bar{x}$. Define G_0 and $G_0^{'}$ analogously to F_0 and $F_0^{'}$, i.e.

$$G_0 = \left\{ d \middle| d \neq 0, \nabla g_i(\bar{x})^t d < 0, i \in I \right\}$$
$$G_0^{'} = \left\{ d \middle| d \neq 0, \nabla g_i(\bar{x})^t d \leq 0, i \in I \right\}$$

Then

$$G_0 \subset D \subset G_0^{'}$$

Proof. Let $d \in G_0$. Since $\bar{x} \in X$ and X is open there exists $\delta_1 > 0$ such that $\bar{x} + \lambda d \in X$ for $\lambda \in (0, \delta_1)$. Also, since $g_{M-I}(\bar{x}) < 0$ and g_{M-I} are continuous at $\bar{x}$ there exists $\delta_2 > 0$ such that

$$g_{M-I}(\bar{x} + \lambda d) < 0 \text{ for } \lambda \in (0, \delta_2)$$

Furthermore since $d \in G_0$, $\nabla g_I(\bar{x})^t d < 0$ and so by the increasing direction theorem there exists $\delta_3 > 0$ such that

$$g_I(\bar{x} + \lambda d) < g_I(\bar{x}) = 0 \text{ for } \lambda \in (0, \delta_3)$$

Let $\delta = \min\{\delta_1, \delta_2, \delta_3\} > 0$ and it's clear that $x = \bar{x} + \lambda d \in S$ for $\lambda \in (0, \delta)$ (why? all constraints are satisfied). Thus $d \in D$, where D is the cone of feasible directions of the feasible region at $\bar{x}$. Thus $G_0 \subset D$. Similarly if $d \in G_0^{'}$ if $\nabla g_i(\bar{x})^t d > 0$ for any $i \in I$ then $g_i(\bar{x} + \lambda d) > g_i(\bar{x}) = 0$ for λ sufficiently small, contradicting that $d \in D$, and thus $D \subset G_0^{'}$. □

Finer results in the case of convexity/concavity

Lemma. *If g_I are strictly convex at $\bar{x}$ then $D = G_0$ and if g_I are concave then $D = G_0^{'}$.*

Proof. Suppose g_I are strictly convex at $\bar{x}$ and let $d \in D$. If $d \notin G_0$ then $\nabla g_i(\bar{x})^t d \geq 0$ for some $i \in I$ we would have $g_i(\bar{x} + \lambda d) > g_i(\bar{x}) = 0$ for all $\lambda > 0$ condtradicting $d \in D$. Hence in this case $D = G_0$. Now suppose g_I are concave at $\bar{x}$ and let $d \in G_0^{'}$. Then $g_I(\bar{x} + \lambda d) \leq g_I(\bar{x})$ for all $\lambda \geq 0$ and by continuity of g_{M-I} and since X is open we obtain $\bar{x} + \lambda d \in S$ for all λ sufficiently small and hence $d \in D$. Hence $G_0^{'} \subset D \subset G_0^{'}$. □

Theorem. *Consider*

$$\min\left\{f(x)\,\middle|\,g_i(x) \leq 0, i \in M = \{1, \ldots, m\}\right\}$$

Let $\bar{x}$ be feasible and I as before. Assume that f and g_I are differentiable at $\bar{x}$ and g_{M-I} are continuous at $\bar{x}$. Then

(1) If $\bar{x}$ is locally optimal then $F_0 \cap G_0 = \emptyset$.
(2) If $F_0 \cap G_0 = \emptyset$ and f is convex, g_I strictly convex in some neighborhood of $\bar{x}$, then $\bar{x}$ is a local minimum.

Proof. Let $\bar{x}$ be locally optimal then $F_0 \cap D = \emptyset$ by two theorems above and by $G_0 \subset D$ we have $F_0 \cap G_0 = \emptyset$. Conversely suppose $F_0 \cap G_0 = \emptyset$ and f, g_i are as specified. Redefining the feasible region as those x which satisfy the binding constraints we have that $G_0 = D$ by above lemma about strict convexity and hence $F_0 \cap D = \emptyset$. Further since the level sets $g_I(x) \leq 0$ are convex over some $\mathcal{N}_\epsilon(\bar{x})$ it follows that $S \cap N_\epsilon(\bar{x})$ is convex (S is the feasible region spanned by feasible directions). Since $F_0 \cap D = \emptyset$ and f is convex at $\bar{x}$ from the converse of improving directions set theorem that $\bar{x}$ is a local minimum. Throwing the nonbinding constraints back in doesn't change anything. □

Now to finally turn the geometric constraint $F_0 \cap G_0 = \emptyset$ into something useful:

Theorem. *Fritz-John "necessary" conditions. Let X be nonempty open. Consider*

$$\min\left\{f(x)\,\middle|\,x \in X, g_i(x) \leq 0, i \in M = \{1, \ldots, m\}\right\}$$

Let $\bar{x}$ be feasible and $g_I(\bar{x}) = 0$. Furthermore suppose that f and g_I are differentiable and g_{M-I} are continuous at $\bar{x}$. If $\bar{x}$ is locally optimal then there exist Lagrange multipliers u_0, u_I such that

$$\begin{aligned} u_0 \nabla f(\bar{x}) + \sum_{i=I} u_i \nabla g_i(\bar{x}) &= 0 \\ u_0, u_I &\geq 0 \\ (u_0, u_I) &\neq (0, 0) \end{aligned}$$

Alternatively this can be rewritten as

$$\begin{aligned} u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) &= 0 \\ u_I g_I(\bar{x}) &= 0 \\ u_0, u_I &\geq 0 \\ (u_0, u) &\neq (0, 0) \end{aligned}$$

Proof. Since $\bar{x}$ is locally optimal $F_0 \cap G_0 = \emptyset$, i.e. there exists no d such that $\nabla f(\bar{x})^t d < 0$ and $\nabla g_I(\bar{x})^t d < 0$. Let

$$\nabla L = \begin{bmatrix} \nabla f(\bar{x})^t \\ \nabla g_i(\bar{x})^t \end{bmatrix}$$

Then this is equivalent to

$$\left\{ d \middle| \nabla L d < 0 \right\} = \emptyset$$

By Gordan's theorem there exists non-zero $p \geq 0$ such that $(\nabla L)^t p = 0$. Denoted the components of $p = (u_0, u_I)$ we have

$$\begin{aligned} u_0 \nabla f(\bar{x}) + \sum_{i=I} u_i \nabla g_i(\bar{x}) &= 0 \\ u_0, u_I &\geq 0 \\ (u_0, u_I) &\neq (0,0) \end{aligned}$$

Complementary slackeness gives us

$$\begin{aligned} u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) &= 0 \\ u_M g_M(\bar{x}) &= 0 \\ u_0, u &\geq 0 \\ (u_0, u) &\neq (0,0) \end{aligned}$$

□

Example. Consider the problem

$$S = \left\{ (x,y) \in \mathbb{R}^2 \middle| \begin{array}{c} g_1(x,y) = x + y^2 - 1 \leq 0 \\ g_2(x,y) = 4x + 3y - 3 \leq 0 \\ g_3(x,y) = -x \leq 0 \end{array} \right\}$$

The figure is 7.1.

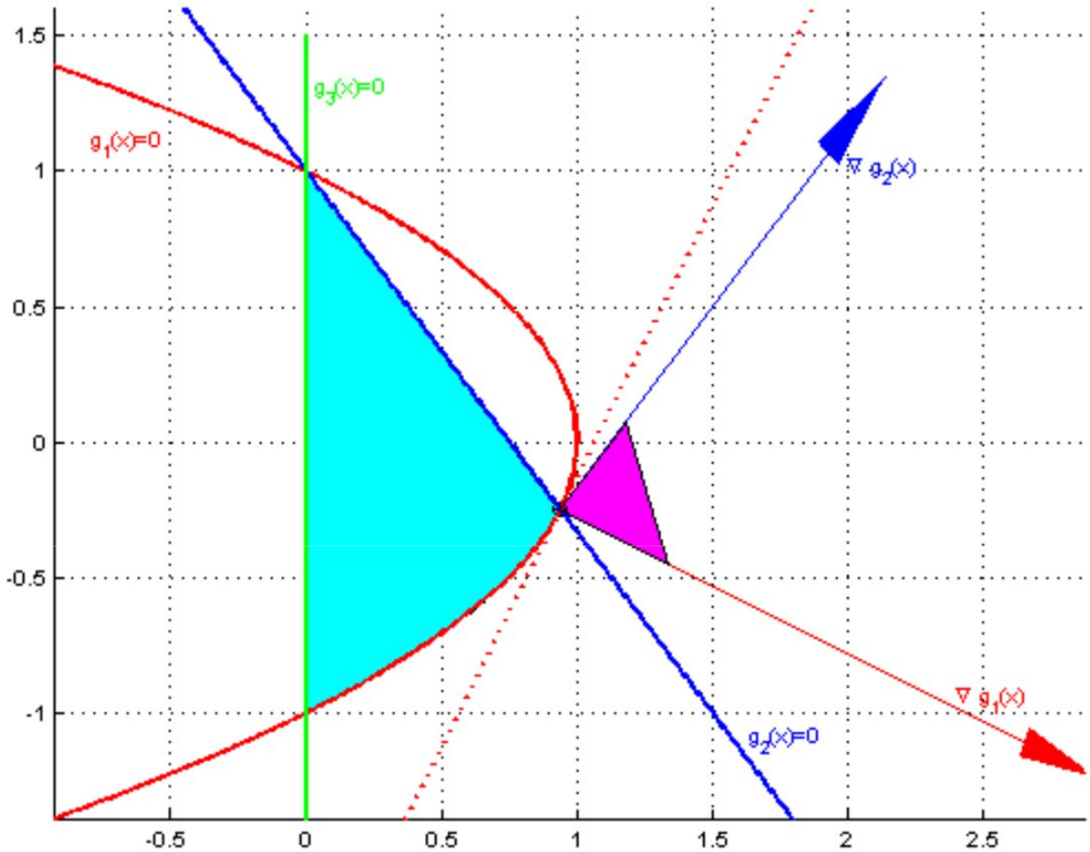

FIGURE 7.1. Inequality constrained feasible region

Note a trick: we can make any feasible $\bar{x}$ point to some minimization problem a Fritz-John point of a related system by adding the constraints $\|x-\bar{x}\|^2 \geq 0$.

Theorem. *KKT conditions. Suppose all the conditions of Fritz-John are met and $\nabla g_I(\bar{x})$ are linearly independent. Then*

$$\begin{aligned} \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) &= 0 \\ u_i g_i(\bar{x}) &= 0 \\ u_i &\geq 0 \end{aligned}$$

Proof. By Fritz-John there exist

$$\begin{aligned} u_0 \nabla f(\bar{x}) + \sum_{i=I} u_i \nabla g_i(\bar{x}) &= 0 \\ u_0, u_I &\geq 0 \\ (u_0, u_I) &\neq (0,0) \end{aligned}$$

and $u_0 > 0$ since otherwise $\{\nabla g_i(\bar{x})\}$ would be linearly dependent. Hence dividing by u_0 we get the conclusion. □

Connection of KKT to first-order LP approximations to NLPs:

Theorem. *Let X be nonempty open. Consider*

$$\min \left\{ f(x) \,\middle|\, x \in X, g_i(x) \leq 0, i \in M = \{1, \ldots, m\} \right\}$$

with f, g_i differentiable. Let $\bar{x}$ be feasible and $g_I(\bar{x}) = 0$. Let $F_0, G_0^{'} = \left\{ d \neq 0 \middle| \nabla g_I(\bar{x})^t d \leq 0 \right\}, G^{'} = G^{'} \cup \{0\}$. Then $\bar{x}$ is KKT iff $F_0 \cap G^{'} = \emptyset \iff F_0 \cap G_0^{'} = \emptyset$. Further $\bar{x}$ is KKT iff

$$\bar{x} = \min_x \left\{ f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) \middle| g_I(\bar{x}) + \nabla g_I(\bar{x})^t (x - \bar{x}) \leq 0 \right\}$$

Theorem. *Sufficient conditions for KKT point to be locally optimal. If $\bar{x}$ is a KKT solution and there exists a neighborhood $\mathcal{N}_\epsilon(\bar{x})$ such that f is convex over $\mathcal{N}_\epsilon(\bar{x}) \cap S$, the g_I are differentiable and convex over $\mathcal{N}_\epsilon(\bar{x}) \cap S$, then $\bar{x}$ is a local minimum.*

Theorem. *Sufficient conditions for KKT point to be globally optimal. If*

$$\min \left\{ f(x) \middle| x \in X, g_i(x) \leq 0, i \in M = \{1, \ldots, m\} \right\}$$

is convex (i.e. f, g_i are convex everywhere) and $\bar{x}$ is a KKT point. Then $\bar{x}$ is globally optimal.

Note that not every every globally optimal point of a convex program is a KKT point. Also if the constraints are equalities then the multipliers are free (not necessarily positive) and if the constraints are greater than then the multipliers are negative.

7.1. **Second order conditions for local optimality.** Consider the problem

$$\begin{aligned} &\min f(x) \\ &\text{s.t.} g_I(x) \leq 0 \\ &\quad h_J(x) = 0 \end{aligned}$$

Definition. The *Lagrangian* is

$$L(x, u, v) = f(x) + \sum_{i \in I} u_i g_i(x) + \sum_{j \in J} v_j h_j(x)$$

Theorem. *Sufficient conditions. Suppose $\bar{x}$ is a KKT point with Lagrange multipliers $\bar{u}, \bar{v}$. Then*

(1) If $\nabla L^2(x) = \nabla L^2(x, \bar{u}, \bar{v})$ is psd for all $x \in S$, then $\bar{x}$ is a global minimum.
(2) If $\nabla L^2(x) = \nabla L^2(x, \bar{u}, \bar{v})$ is psd for all $x \in \mathcal{N}_\epsilon(\bar{x}) \cap S$, then $\bar{x}$ is a local minimum.
(3) If $\nabla L^2(x) = \nabla L^2(x, \bar{u}, \bar{v})$ is pd then $\bar{x}$ is a strict local minimum.

Theorem. *Necessary conditions. Suppose $\bar{x}$ is a local minimum Lagrange multipliers $\bar{u}, \bar{v}$ and $\nabla g_I(\bar{x})$ (where I is the set of binding constraints) are linearly independent and $\nabla h_J(\bar{x})$ are linearly independent. Then $\bar{x}$ is a KKT point having Lagrange multipliers $\bar{u} \geq 0$ and $\nabla^2 L(\bar{x})$ is psd over*

$$C = \left\{ d \neq 0 \middle| \begin{array}{ll} \nabla g_{I^+}(\bar{x})^t d = 0 & I^+ = \left\{ i \in I \middle| \bar{u}_i \geq 0 \right\} \\ \nabla g_{I^0}(\bar{x})^t d \leq 0 & I^0 = \left\{ i \in I \middle| \bar{u}_i = 0 \right\} \\ \nabla h_J(\bar{x})^t d = 0 & \end{array} \right\}$$

Theorem. *Let $\bar{x}$ be a KKT point. If $\nabla^2 L(\bar{x})$ is pd over C then $\bar{x}$ is a strict local minimum.*

In summary

8. Lagrangian Duality

Definition. Consider the following related pair of optimization problems

$$\begin{aligned} z^P &= \min \left\{ f(x) \middle| x \in X \right\} \\ z^R &= \min \left\{ g(x) \middle| x \in Y \right\} \end{aligned}$$

R is a *relaxation* of P if $f(x) \geq g(x)$ for all $x \in X$ and $X \subset Y$.

Theorem. *Let R be a relaxation of P*

(1) *If R is infeasible then so is P.*
(2) *If P and R have optimal solutions then $z^P \geq z^R$*
(3) *If x^R is an optimal solution to R such that $x^R \in X$ and $g\left(x^R\right) = f\left(x^R\right)$, then x^R is an optimal solution for P.*

Definition. Consider the primal optimization problem

$$\begin{aligned}
&\min f(x) \\
&\text{s.t.} g_i(x) \leq 0 \text{ for all } i = 1, \ldots, m \\
&\quad h_j(x) = 0 \text{ for all } j = 1, \ldots, n \\
&\quad x \in X
\end{aligned}$$

The Lagrangian dual is

$$\begin{aligned}
&\max \theta(u, v) \\
&\text{s.t.} u \in \mathbb{R}_+^m, v \in \mathbb{R}^n
\end{aligned}$$

where

$$\theta(u, v) = \min_x L(x, u, v)$$

where

$$L(x, u, v) = f(x) + \sum_{i=1}^m u_i g_i(x) + \sum_{j=1}^n v_j h_j(x)$$

Example. Consider the problem

$$\begin{aligned}
&\min x_1^2 + x_2^2 \\
&\text{s.t.} - x_1 - x_2 + 4 \leq 0 \\
&\quad x_1, x_2 \geq 0
\end{aligned}$$

"Dualize" the first constraint:

$$L(u, x_1, x_2) = x_1^2 + x_2^2 + u(-x_1 - x_2 + 4)$$

Then the dual function is

$$\begin{aligned}
\theta(u) &= \min_{x_1, x_2} L(u, x_1, x_2) \\
&= \min_{x_1, x_2} \left(x_1^2 - u x_1 + x_2^2 - u x_2 + 4u\right) \\
&= 4u + \min_{x_1 \geq 0} \left(x_1^2 - u x_1\right) + \min_{x_2 \geq 0} \left(x_2^2 - u x_2\right) \\
&= 4u + \left\{ \begin{matrix} -\frac{u^2}{4} \text{ if } u \geq 0 \\ 0 \text{ if } u < 0 \end{matrix} \right\} + \left\{ \begin{matrix} -\frac{u^2}{4} \text{ if } u \geq 0 \\ 0 \text{ if } u < 0 \end{matrix} \right\} \\
&= \left\{ \begin{matrix} 4u - \frac{u^2}{2} \text{ if } u \geq 0 \\ 4u \text{ if } u < 0 \end{matrix} \right\}
\end{aligned}$$

8.1. **A general procedure for deriving duals.** Procedure:

(1) Build the Lagrangian function (choose the appropriate constraints to "dualize").
(2) For fixed Langrange multipliers do the minimization (or maximization).
(3) Simplify the description of the Lagrangian.
(4) Express the dual.

8.2. **Deriving duals.**

Example. The Langrangian dual of the QP

$$\min\left\{d^tx+\frac{1}{2}x^tHx\middle|Ax\preceq b\right\}$$

where H is pd and symmetric is

$$\max\left\{\frac{1}{2}u^tDu+u^tC-\frac{1}{2}d^tH^{-1}d\middle|u\succeq 0\right\}$$

where $D=-AH^{-1}A^t, C=-b-AH^{-1}d$. Why?

(1) Build the Lagrangian

$$\begin{aligned}L(u)&=\min_{x\in\mathbb{R}^n}\left\{d^tx+\frac{1}{2}x^tHx+u^t(Ax-b)\right\}\\&=\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}x^tHx+\left(d^t+u^tA\right)x-u^tb\right\}\\&=-u^tb+\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}x^tHx+\left(d^t+u^tA\right)x\right\}\end{aligned}$$

(2) For a fixed u the inner minimization is convex and unconstrained, so we can use first order conditions: solve for x

$$Hx+\left(d^t+u^tA\right)=0$$

i.e. $x^*=-H^{-1}\left(d+A^tu\right)$

(3) Simplifying $L(u)$ we get

$$L(u)=-\frac{1}{2}d^tH^{-1}d+u^t\left(-b-AH^{-1}d\right)-\frac{1}{2}u^tAH^{-1}A^tu$$

(4) Express the dual

$$\max\left\{\frac{1}{2}u^tDu+u^tC-\frac{1}{2}d^tH^{-1}d\middle|u\succeq 0\right\}$$

Note that $u\succeq 0$ because $Ax-b\preceq 0$ and hence us should be positive in order for $u(Ax-b)$ to be a relaxation.

Definition. Given a closed convex cone $K\subset\mathbb{R}^n$, the *dual cone* K^* of K

$$K^*=\left\{y\in\mathbb{R}^n\middle|y^tx\geq 0,\forall x\in K\right\}$$

I.e. all vectors that make an acute angle with every vector in K.

Example. Some cones are "self-dual"

(1) $\left(\mathbb{R}^n_+\right)^*=\mathbb{R}^n_+$
(2) $(\mathcal{C}_n)^*=\mathcal{C}_n$ where $\mathcal{C}_n$ is the n dimensional second order cone.
(3) $\left(S^{n\times n}_+\right)^*=S^{n\times n}_+$ where $S^{n\times n}_+$ is the cone of $n\times n$ psd matrices.

Theorem. *If K is a nonempty closed convex cone then so is K^*.*

Example. The Lagrangian dual of the conic LP

$$\min\left\{c^tx\middle|Ax=b,x\in K\right\}$$

is

$$\max\left\{b^tv\middle|A^tv+s=c,s\in K^*\right\}$$

Why?

(1) Build the Lagrangian dual

$$\begin{aligned}L(v)&=\min_{x\in K}\ c^tx+v^t(b-Ax)\\&=v^tb+\min_{x\in K}\ \left(c^t-v^tA\right)x\end{aligned}$$

(2) For a fixed u this is optimizing a linear function over a cone K, therefore the problem will be bounded iff

$$c - A^t v \in K^*$$

(otherwise we could just "runaway"). Similarly as in the straightforward LP case we have that if there is an optimal solution it's at 0.

(3) Simplify the Lagrangian

$$L(v) = \begin{cases} b^t v & \text{if } c - A^t v \in K^* \\ -\infty & \text{if } c - A^t v \notin K^* \end{cases}$$

Theorem. *The dual of the dual of the conic LP*

$$\min \left\{ c^t x \middle| Ax = b, x \in K \right\}$$

is

$$\min \left\{ c^t x \middle| Ax = b, x \in (K^*)^* \right\}$$

Theorem. *If K is a nonempty closed convex cone then $(K^*)^* = K$.*

8.3. **The strength of duals.**

Theorem. *Weak duality. Let x be a feasible solution to a primal problem and (u, v) be a feasible solution to the dual. Then $f(x) \geq L(u, v)$.*

Corollary. *The optimal solution to the primal is greater or equal to the optimal solution to the dual.*

Corollary. *If $\bar{x}$ and $(\bar{u}, \bar{v})$ are such that $f(\bar{x}) = L(\bar{u}, \bar{v})$ then $\bar{x}$ solves the primal and $(\bar{u}, \bar{v})$ solves the dual.*

Corollary. *If the primal is unbounded then $L(u, v) = -\infty$ for all u, v.*

Corollary. *If the dual is unbounded then the primal is infeasible.*

Definition. When the optimal solution to the primal and the optimal solution to the dual differ there's a duality gap.

Lemma. *Let X be a nonempty convex subset of $\mathbb{R}^n$, $\alpha : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ convex functions and $h = Ax - b$.*

$$\begin{aligned} \{\alpha < 0, g(x) \leq 0, h(x) = 0\} &= \emptyset \\ &\Rightarrow \\ \left\{ u_0 \alpha(x) + u^t g(x) + v^t h(x) \geq 0, \forall x \in X, (u_0, u) \geq 0, (u_0, u, v) \neq 0 \right\} &\neq \emptyset \end{aligned}$$

The converse holds if $u_0 > 0$.

Theorem. *Strong Duality. There is no duality gap if there exists $\hat{x} \in X$ such that $g(\hat{x}) < 0, h(\hat{x}) = 0, 0 \in int(h(X))$*

Definition. A *saddle point of the Lagrangian is $(\bar{x}, \bar{u}, \bar{v})$ such that*

$$L(\bar{x}, u, v) \leq L(\bar{x}, \bar{u}, \bar{v}) \leq L(x, \bar{u}, \bar{v})$$

Theorem. *A solution $(\bar{u}, \bar{v}, \bar{x})$ with $\bar{u} \geq 0$ is a saddle point for the Lagrangian iff*

(1) $L(\bar{u}, \bar{v}, \bar{x}) = \min \{L(\bar{u}, \bar{v}, x)\}$
(2) $g(\bar{x}) \leq 0$ *and* $h(\bar{x}) = 0$
(3) $\bar{u}^t g(\bar{x}) = 0$
(4) *$\bar{x}$ and $(\bar{u}, \bar{v})$ are optimal for the primal problem and the dual problem respectively and there's no duality gap.*

Fact: the dual function is always concave and it's differentiable when something something something is unique.

9. Sundry things

Theorem. *Suppose H is symmetric and psd and $x^t H x = 0$ then $Hx = 0$.*

Proof. Let QDQ^t be the eigenvalue decomposition of H. Since H is psd $D = \sqrt{D}\sqrt{D}$. Then

$$\begin{aligned} x^t H x &= 0 \\ &\Longleftrightarrow \\ x^t Q\sqrt{D}\sqrt{D}Q^t x &= 0 \\ &\Longleftrightarrow \\ \left\|\sqrt{D}Q^t x\right\| &= 0 \\ &\Longleftrightarrow \\ \sqrt{D}Q^t x &= 0 \\ &\Longleftrightarrow \\ Q\sqrt{D}\sqrt{D}Q^t x &= 0 \\ &\Longleftrightarrow \\ Hx &= 0 \end{aligned}$$

□

Example. Consider

$$\min_x \ \|Ax - b\|^2$$

If b is in the column space then this is just solving $Ax = b$. If the system is underdetermined then this finds the min norm solution. If b is not in the column space of A then this find the projection b onto the column space of A. If the norm is the 2norm then the objective

$$f(x) = b^t b - 2x^t A^t b + x^t A^t A x$$

and the first order necessary condition $\nabla f(x) = 0$ is $A^t A x = A^t b$. The Hessian of f is $A^t A$, which is psd ($x^t A^t A x = \|Ax\|$). Therefore f is convex and hence the necessary condition is also sufficient. If A is full rank then $A^t A$ is invertible. If A is not full rank but is fullrow rank then additional criteria can be used to invert: solve the problem

$$\begin{aligned} &\min_x \ x^t x \\ &\text{s.t.} Ax = b \end{aligned}$$

Using again KKT you get $x = A^t \left(AA^t\right)^{-1} b$.

Example. Consider minimizing f subject to $g_i(x) \le 0$. Verifying $\bar{x}$ is a KKT point is equivalent to finding a solution $u \ge 0$

$$\nabla f(\bar{x}) + \sum_{i \in I} u_i g_i(\bar{x}) = 0$$

where I is the set of binding constraints. Restated this is

$$A^t u = c, \ u \ge 0$$

where $A^t = [\nabla g_i(\bar{x}), i \in I]$ and $c = -\nabla f(\bar{x})$. This is equivalent to the following LP

$$\begin{aligned} &\min_y \ e^t y \\ &\text{s.t.} A^t u \pm y = c \\ &\quad u \ge 0, y \ge 0 \end{aligned}$$

having objective value 0, where $e = (1, \ldots, 1)$ and y is slack that has the same signs as c.

Example. Prove the AM-GM inequality

$$\frac{1}{n}\sum_{j=1}^{n} x_j \geq \left(\prod_{j=1}^{n} x_j\right)^{1/n}$$

using KKT. Consider the problem

$$\begin{aligned} \min \ & \sum_{j=1}^{n} x_j \\ \text{s.t.} & \prod_{j=1}^{n} x_j = b \\ & x_j \geq 0 \end{aligned}$$

where $b > 0$. Therefore $x_j > 0$ and the only active constraint is the equality constraint and by linear independence the KKT conditions are necessary. The KKT system is

$$\begin{aligned} 1 + v\prod_{i \neq j}^{n} x_i = 0 \\ \prod_{j=1}^{n} x_j = b \end{aligned}$$

Primal feasible constraints. Multiplying the jth equation by x_j we get

$$x_j + vb = 0$$

and therefore

$$\sum_{j=1}^{n} x_j + nbv = 0$$

which gives the unique value for the Lagrange multiplier of

$$v = -\frac{1}{nb}\sum_{j=1}^{n} x_j$$

Substituting into each KKT equation we get

$$x_j = \frac{1}{n}\sum_{j=1}^{n} x_j = \bar{x}$$

and which shows the values of all of the $\bar{x} \coloneqq x_j$ are all identical

$$\prod_{j=1}^{n} x_j = (\bar{x})^n = b \Rightarrow \bar{x} = b^{1/n}$$

and since the KKT conditions are necessary this gives the unique optimum? I.e.

$$\frac{1}{n}\sum_{j=1}^{n} \bar{x} = b^{1/n}$$

is the optimal objective. Thus

$$\frac{1}{n}\sum_{j=1}^{n} x_j \geq \min\left\{\frac{1}{n}\sum_{j=1}^{n} x_j \middle| \prod_{j=1}^{n} x_j = b\right\} = b^{1/n} = \left(\prod_{j=1}^{n} x_j\right)^{1/n}$$

10. Summary

Fact 1. ***Min LP at extreme points***

(1) *For a Minimization LP there's a solution iff* $c^t d_j \geq 0$ *for all extreme directions* d *and objective* c*. Why? Intuitively no extreme direction (if any exist at all) should be aligned with* $-c$ *(because otherwise you could minimize forever).*
(2) *The solution is at a vertex of the polytope (Minkowski's representation theorem since all the* $c^t d_j \geq 0$*).*

Fact 2. ***Convex iff convex for every direction****: a multivariate function* f *is convex iff it's convex on every line, i.e. iff*

$$F_{\bar{x},d}(\lambda) := f(\bar{x} + \lambda d)$$

is a convex function of λ *for all* $\bar{x}, d$*.*

Fact 3. ***Alpha level sets of convex functions are convex****: for convex* S *and convex* f*, the set*

$$S_\alpha := \left\{x \in S \middle| f(x) \leq \alpha\right\}$$

is convex. Easy to prove just from the definitions.

Fact 4. ***Epigraphs****. A function is convex iff its epigraph is convex.*

Fact 5. ***Convex functions are continuous on their interiors****. How to remember this? If it weren't continuous (had a jump discontinuity) it wouldn't be convex.*

Fact 6. ***Seperating hyperplane theorem****. For convex* S *and all* $\bar{x} \in \partial S$ *there exists* $p \neq 0$ *such that* $p^t(x - \bar{x}) \leq 0$*.*

Fact 7. ***Subgradients****. Convex functions over convex sets have supporting hyperplanes, i.e.*

$$f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x})$$

Note that epi(f) *is convex and* $(\bar{x}, f(\bar{x}))$ *belongs to its boundary. Therefore by separation theorem there exists* $(\xi_0, \mu) \in \mathbb{R}^n \times \mathbb{R}$*. For all* $(\mathbf{x}, y) \in$ *epi*(f)*.*

$$\xi_0^t(x - \bar{x}) + \mu(y - f(\bar{x})) \leq 0$$

Note that $\mu \leq 0$ *because otherwise take* y *large enough and the inequality would be violated.*

Fact 8. ***Converse about subgradients****. If there exists a supporting hyperplane for the epigraph of* f *at every point* $x \in S$ *a convex set then* f *is convex.*

Fact 9. ***Farkas' Lemma****. Only one of these systems has a solution*

$$\left\{x \middle| Ax \preceq 0, c^t x > 0\right\}$$
$$\left\{y \middle| A^t y = c, y \succeq 0\right\}$$

Fact 10. ***Extreme points****. For* $A \in \mathbb{R}^{m\times n}$*,* m *constraints,* n *variables.* x *is an extreme point iff* A *can be partitioned*

$$A = [B, N]$$

where $B \in \mathbb{R}^{m\times m}$ *and invertible and* $Ax = Bx_B + Nx_N$ *where* $x_N = 0$ *and* $x_B = B^{-1}b$*.*

Fact 11. ***Simplex method****.* ***For a maximization problem****: write down the tableua with the top row being* $z - c^t x$*. If any of the coefficients are negative then you can increase the objective pivoting on that column. Choose which row to pivot by using the "smallest ratio" test.*

Fact 12. ***Extreme directions****. The extreme directions of a polyhedron are the extreme points of*

$$D = \left\{d \middle| Ad \preceq 0, d \succeq 0, 1^t d = 1\right\}$$

Fact 13. ***Directional derivatives****. Convex functions have all directional derivatives.*

Fact 14. ***Subdifferential of differentiable function***. *Using the supporting hyperplane theorem for convex epigraphs we have that the subgradient of a differentiable function is unique, i.e.* $\nabla f(\bar{x})$. *Therefore for differentiable* f

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x})$$

Fact 15. ***Another weird convexity condition***. *If* f *is differentiable then* f *is convex iff*

$$(\nabla f(x_2) - \nabla f(x_1))^t (x_2 - x_1) \geq 0$$

Fact 16. ***Twice differentiable***. *If* f *is twice differentiable then* f *is convex iff* $\nabla^2 f(\bar{x})$ *is psd everywhere. If* $\nabla^2 f(\bar{x})$ *is pd everywhere then* f *is strictly convex.*

Fact 17. ***Optimality conditions***. *If* f *is convex and* $\bar{x}$ *is locally optimal then (**the obvious convex thing**).*

(1) $\bar{x}$ *is globally optimal.*
(2) *If furthermore* $\bar{x}$ *is strictly locally optimal then*
  (a) $\bar{x}$ *is unique.*
  (b) $\bar{x}$ *is isolated.*

Fact 18. ***More optimality conditions***. *For minimization problem if* f *is convex then* $\bar{x}$ *is globally optimal iff there exists a subgradient* ξ *such that* $\xi^t (x - \bar{x}) \geq 0$.

Corollary. $\bar{x}$ *is globally optimal iff* $0 \in \partial f(\bar{x})$.

Fact 19. *For a maximization problem if* $\bar{x}$ *is lcoally optimal then* $\xi^t (x - \bar{x}) \leq 0$. *If* f *is convex then the maxima occur at the boundary (obvious type thing).*

Fact 20. *For an unconstrained minimization problem if* d *is such that* $\nabla f(\bar{x})^t d < 0$ *then* d *is non-improving. If* $\bar{x}$ *is a local optimum then* $\nabla f(\bar{x}) = 0$ *and* $\nabla^2 f(\bar{x})$ *is psd (these are necessary and sufficient in the unconstrained case). Think calculus. For strict local minima* $\nabla^2 f(\bar{x})$ *must be pd.*

Fact 21. *Let*

$$F = \left\{ d \middle| f(\bar{x} + \lambda d) < f(\bar{x}) \text{ s.t. } \exists_\delta \forall_\lambda \ \lambda \in (0, \delta) \right\}$$

be the cone of improving directions and

$$D = \left\{ d \middle| \bar{x} + \lambda d \in S \text{ s.t. } \exists_\delta \forall_\lambda \ \lambda \in (0, \delta) \right\}$$

be the set of feasible directions

$$F_0 = \left\{ d \middle| \nabla f(\bar{x})^t d < 0 \right\}$$

then $F_0 \subseteq F$.

Fact 22. ***Improving directions intersect*** F_0. *If* $\bar{x}$ *is a local optimum then* $F_0 \cap D = \emptyset$.

Fact 23. ***Converse***. *If* f *is convex and* $F_0 \cap D = \emptyset$ *and there exists a neighborhood such that every direction is feasible then* $\bar{x}$ *is a local optimum.*

Fact 24. *Let* $F_0^{'} = \left\{ d \neq 0 \middle| \nabla f(\bar{x})^t d \leq 0 \right\}$. *If* F *is convex then* $F_0 = F$. *If* f *is concave then* $F = F_0^{'}$.

Fact 25. *Let* $G_0 = \left\{ d \middle| \nabla g_i(\bar{x})^t d < 0 \right\}$ *and* $G_0^{'} = \left\{ d \middle| \nabla g_i(\bar{x})^t d \leq 0 \right\}$. *Let* I *be the subset of constraints that are binding. If* g_I *are strictly convex then* $G_0 = D$ *and if* g_I *are strictly concave then* $D = G_0^{'}$.

Fact 26. ***Geometric optimality necessary conditions***. *If* $\bar{x}$ *is feasible and locally optimal and* f *and* g_I *are differentiable and nonbinding* g_i *are continuous then* $F_0 \cap G_0 = \emptyset$.

Fact 27. ***Geometric optimality sufficient conditions***. *If* $F_0 \cap G_0 = \emptyset$ *and* f *is convex and* g_I *binding strictly convex (*$F = F_0$ *and* $D = G_0$*) then* $\bar{x}$ *is locally optimal.*

Fact 28. ***Algebraic optimality necessary conditions***. ***FJ conditions***. *If $\bar{x}$ is locally optimal and every differentiable and continuous then there exists u_0, u such that*

$$u_0 \nabla f(\bar{x}) + \sum_i u_i \nabla g_i(\bar{x}) = 0$$
$$u_i g_i(\bar{x}) = 0$$
$$u_0, u \succeq 0$$
$$(u_0, u) \neq 0$$

The second row is called the complementary slackness condition, i.e. binding constraints for that to always be zero and otherwise the $u_i = 0$.

Fact 29. ***KKT "necessary" conditions***. *KKT conditions are the same as FJ except the binding $\nabla g_i(\bar{x})$ should be zero.*

Fact 30. ***KKT sufficient condtions***. *If g_i, f_i are convex and $\bar{x}$ is a KKT point then $\bar{x}$ is globally optimal.*

Fact 31. ***Reiterating***.

$$\text{Local optimality} \rightarrow F_0 \cap D = \emptyset \underset{\text{Gordan's theorem}}{\rightarrow} \rightarrow FJ$$

Let T be the cone of tangents. Then local optimality implies $F_0 \cap T = \emptyset$. Under constraint qualifications $T = G_0^{'}$

$$\text{Local optimality} \rightarrow F_0 \cap G_0^{'} = \emptyset \underset{\text{Farkas' lemma}}{\rightarrow} \rightarrow KKT$$

Fact 32. ***Weak duality***. *Let $L(x, u, v)$ be the Lagrangian. Then*

$$f(x) \geq L(x, u, v)$$

Fact 33. *The dual function, for a min problem is*

$$\phi(u, v) = f(x) + L(u, v) = \min_x \left(c^t x + u^t (b - Ax) + v^t (e - Dx)\right) = u^t b + v^t e + \min_x \left(\left(c^t - A - D\right) x\right)$$

and the dual problem is

$$\max_{u,v} \phi(u, v)$$
$$s.t. u \geq 0$$
$$v \text{ free}$$

If the dual primal are equal then the solution is the optimum, i.e. there's no duality gap. If there's a saddle point

$$L(\bar{x}, u, v) \leq L(\bar{x}, \bar{u}, \bar{v}) \leq L(x, \bar{u}, \bar{v})$$

in Lagrangian then there's no duality gap.

Fact 34. ***Abadie's Constraints qualification***. *Tangent cone = Linearized cone of constraints.*

Fact 35. ***Slater's condition***. *This is for there to be no duality gap. There exists a point in the interior of the feasible region, i.e. $g_i(x) < 0$.*