AMSC 607 / CMSC 764 Advanced Numerical Optimization
Fall 2003
UNIT 3: Constrained Optimization
PART 1: Characterizing a solution
Dianne P. O'Leary

Fundamentals for constrained optimization

- Characterizing a solution
- Duality

Our approach: Always try to reduce the problem to one with a known solution.

Reference: N&S, Chapter 14

Our problem

$$\min_x f(x)$$

$$\begin{aligned} c_i(x) &= 0\,, \quad i \in \mathcal{E} \\ c_i(x) &\geq 0\,, \quad i \in \mathcal{I} \end{aligned}$$

where f and c_i are $\mathcal{C}^2$ functions from $\mathcal{R}^n$ into $\mathcal{R}^1$.

Definition of a solution

We say that x^* is a **solution** to our problem if

- x^* satisfies all of the constraints.
- For some $\epsilon > 0$, if $\|y - x^*\| \leq \epsilon$, and if y satisfies the constraints, then $f(y) \geq f(x^*)$.

In other words, x^* is **feasible** and **locally optimal**.

The plan

We will develop necessary and sufficient **optimality conditions** so that we can recognize solutions and develop algorithms to find solutions.

We do this in several stages.

- Case 1: Linear equality constraints only.
- Case 2: Linear inequality constraints.
- Case 3: General constraints.

Then we will discuss **duality**.

Case 1: Optimality Conditions for Linear equality constraints only

---

Our problem

Reference: Some of this material can be found in N&S Chapter 3.

Our problem:

$$\min_x f(x)$$

$$Ax = b$$

where A is a matrix of dimension $m \times n$.

We also assume a **constraint qualification** or **regularity condition**: assume that A has rank m.

Unquiz:

- What happens if A has rank n?
- What happens if A has rank less than m?

[]

An example

Let

$$\begin{aligned} f(x) &= x_1^2 - 2x_1x_2 + x_2^2 \\ c_1(x) &= x_1 + x_2 - 1 = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 1 \end{aligned}$$

We'll consider two approaches to the problem.

Approach 1: Variable Reduction

If $x_1 + x_2 = 1$, then all feasible points have the form

$$\begin{bmatrix} x_1 \\ 1 - x_1 \end{bmatrix}.$$

Therefore, the possible function values are

$$\begin{aligned} f(x) &= x_1^2 - 2x_1x_2 + x_2^2 \\ &= x_1^2 - 2x_1(1 - x_1) + (1 - x_1)^2 \end{aligned}$$

We now have an **unconstrained minimization problem** involving a function of a single variable, and we know how to solve this!

picture

This is called the **reduced variable method**.

Approach 2: The feasible direction formulation

If $x_1 + x_2 = 1$, then all feasible points have the form

$$x = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

This formulation works because

$$Ax = \begin{bmatrix} 1 & 1 \end{bmatrix} x = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 1$$

and all vectors x that satisfy the constraints have this form.

We obtain this formulation for feasible x by taking a **particular solution**

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and adding on a linear combination of vectors that **span the null space** of the matrix

$$\begin{bmatrix} 1 & 1 \end{bmatrix}.$$

The null space defines the set of **feasible directions**, the directions in which we can step without immediately stepping outside the feasible space.

End example []

What we have accomplished

In general, if our constraints are $Ax = b$, to get feasible directions, we express x as

$$x = \bar{x} + Zv$$

where

- $\bar{x}$ is a particular solution to the equations $Ax = b$ (any one will do),
- the columns of Z form a basis for the nullspace of A (any basis will do),
- v is an arbitrary vector of dimension $(n - m) \times 1$.

Then we have succeeded in reformulating our constrained problem as an unconstrained one:

$$\min_v f(\bar{x} + Zv)$$

Where does Z come from?

N&S, Section 3.3.4

Suppose we have a **QR factorization** of the matrix A^T:

$$A^T = Q\hat{R} \equiv \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R + Q_2 0$$

where

- $Q_1 \in \mathcal{R}^{n\times m}$,
- $Q_2 \in \mathcal{R}^{n\times(n-m)}$,
- $R \in \mathcal{R}^{m\times m}$ is upper triangular,
- $0 \in \mathcal{R}^{(n-m)\times m}$,
- $Q^T Q = I$.

Then

$$Ax = (R^T Q_1^T + 0 Q_2^T)x = R^T Q_1^T x$$

and the columns of Q_2 form a basis for the nullspace of A.

Therefore, to determine Z, we do a QR factorization of A^T and set $Z = Q_2$.

Algorithms for QR factorization: Gram-Schmidt, Givens, Householder, ...

What are the optimality conditions for our reformulated problem?

$$\min_v f(\bar{x} + Zv)$$

Let

$$F(v) = f(\bar{x} + Zv)\,.$$

Then

$$\begin{aligned}
\nabla_v F(v) &= Z^T \nabla_x f(\bar{x} + Zv) = Z^T g(x) \\
\nabla_v^2 F(v) &= Z^T \nabla_x^2 f(\bar{x} + Zv) Z = Z^T H(x) Z
\end{aligned}$$

since $\bar{x} + Zv = x$.

Our theory for unconstrained optimization now gives us **necessary conditions for optimality**:

- **Reduced gradient is zero:** $Z^T \nabla f(x) = 0$.
- **Reduced Hessian** $Z^T \nabla^2 f(x) Z$ **is positive semidefinite**.

We also have **sufficient conditions for optimality**:

- **Reduced gradient is zero:** $Z^T \nabla f(x) = 0$.

- **Reduced Hessian $Z^T \nabla^2 f(x) Z$ is positive definite**.

An alternate approach

Recall what you know, from advanced calculus, about **Lagrange multipliers**: to minimize a function subject to equality constraints, we set up the Lagrange function, with one Lagrange multiplier per constraint, and find a point where its partial derivatives are all zero.

Note: We'll sketch the proof of why this works when we consider nonlinear constraints later in this set of notes.

The Lagrange function for our problem

$$\min_x f(x)$$

$$Ax = b$$

is

$$L(x, \lambda) = f(x) - \lambda^T (Ax - b)\,,$$

and setting the partials to zero yields

$$\begin{array}{rcl} \nabla_x L = & \nabla f(x) - A^T \lambda & = 0\,, \\ -\nabla_\lambda L = & Ax - b & = 0\,. \end{array}$$

These are the **first order necessary conditions for optimality**.

What does this mean geometrically? The solution is characterized by this:

- It satisfies the constraints.
- The gradient of f at x^* is a linear combination of the rows of A, which are the gradients of the constraints.

We can also express this in terms of our QR factorization: $A^T \lambda = g(x)$, means

$$Q_1 R \lambda = g(x)$$

so $g(x)$ is in the range of the columns of Q_1 and this is equivalent to

$$Q_2^T g(x) = 0$$

or, in our earlier notation,

$$Z^T g(x) = 0\,.$$

So we have an **alternate formulation of our first order necessary conditions for optimality**:

$$\begin{aligned} Z^T g(x) &= 0\,, \\ Ax &= b\,. \end{aligned}$$

Three digressions

Digression 1: There are cheaper but less stable alternatives to QR.

The QR factorization gives a very nice basis for the nullspace: its columns are mutually orthogonal and therefore computing with them is stable.

There are alternative approaches.

Option 1: Partitioning

Let

$$A = \begin{bmatrix} B & N \end{bmatrix}$$

where $B \in \mathcal{R}^{m\times m}$ and $N \in \mathcal{R}^{m\times(n-m)}$.

Partition x similarly, with $x_1 \in \mathcal{R}^m$ and $x_2 \in \mathcal{R}^{n-m}$.

Assume that B is nonsingular. (If not, rearrange the columns of A until it is.)

Then $Ax = 0$ if and only if

$$Bx_1 + Nx_2 = 0\,,$$

and this means

$$x_1 + B^{-1}Nx_2 = 0\,,$$

so

$$x_1 = -B^{-1}Nx_2$$

and

$$x = \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix} v\,.$$

Therefore, the columns of

$$\begin{bmatrix} -B^{-1}N \\ I \end{bmatrix}$$

must be a basis for the nullspace of A!

Caution: This basis is sometimes very **ill-conditioned**, and working with it can lead to unnecessary round-off error.

Option 2: Orthogonal projection

Let

$$x = p + q$$

where p is in the nullspace of A and q is in the range of A^T.

Then

$$Ap = 0$$

and q can be expressed as

$$q = A^T\lambda$$

for some vector λ.

Now

$$Ax = A(A^T\lambda)$$

so

$$\lambda = (AA^T)^{-1}Ax.$$

Let's look at

$$\begin{aligned} p &= x - q \\ &= x - A^T(AA^T)^{-1}Ax \\ &= (I - A^T(AA^T)^{-1}A)x \\ &\equiv Px\,. \end{aligned}$$

The matrix P is an **orthogonal projection** that takes x into the null space of A.

Thus we have reduced our problem to an unconstrained one, where $x = x_b + (I - A^T(AA^T)^{-1}A)y$ where x_b is a particular solution to $Ax = b$ and y is any n-vector.

Unquiz: Prove that

1. $P^2 = P$.
2. $P^T = P$.

but note that in general $P^T P \neq I$, so P itself is not an orthogonal matrix. []

The projector P is usually applied using a Cholesky factorization.

Digression 2: the meaning of the Lagrange multipliers

Our optimality conditions:

$$\begin{aligned} g(x^*) - A^T \lambda^* &= 0 \\ Ax^* - b &= 0 \end{aligned}$$

Sensitivity analysis: Suppose we have a point $\hat{x}$ satisfying

$$\|x^* - \hat{x}\| \leq \epsilon$$

and

$$A\hat{x} = b + \delta$$

where ϵ and $\|\delta\|$ are small.

Then Taylor series expansion tells us

$$\begin{aligned} f(\hat{x}) &= f(x^*) + (\hat{x} - x^*)^T g(x^*) + O(\epsilon^2) \\ &= f(x^*) + (\hat{x} - x^*)^T A^T \lambda^* + O(\epsilon^2) \\ &= f(x^*) + \delta^T \lambda^* + O(\epsilon^2)\,. \end{aligned}$$

What this tells us: If we wiggle b_i by δ_i, then we wiggle f by $\delta_i \lambda_i^*$.

Therefore, λ_i^* is the change in f per unit change in b_i. It tells us the **sensitivity** of f to b_i.

Jargon: λ_i is called a **dual variable** or a **shadow price**.

Digression 3

It is important to realize that we do **not** minimize the Lagrangian function

$$L(x, \lambda) = f(x) - \lambda^T (Ax - b) .$$

We find a **saddlepoint** of this function.

So far...

- We have optimality conditions for unconstrained problems.
- We have optimality conditions for linear equality constraints.

Case 2: Optimality conditions for linear inequality constraints

A big "if"

IF we knew

$$\mathcal{W} = \{ i \in \mathcal{I} : c_i(x^*) = 0 \} ,$$

where $c(x^*) = Ax^* - b$, then we could set up the Lagrange multiplier problem and have optimality conditions for our problem.

Let $\bar{W}$ denote the subscripts not in $\mathcal{W}$.

But we don't know the set $\mathcal{W}$ of constraints that are **active** at the solution.

Let's guess!

Suppose we take a guess at the active set. This gives us a set of equations to solve:

$$\begin{aligned} g(x) - A_w^T \lambda_w &= 0, \\ A_w x &= b_w. \end{aligned}$$

Assume that A_w has full row rank. This implies that $\mathcal{W}$ has at most n elements.

Suppose this system has a solution $\hat{x}$, $\hat{\lambda}$. Also suppose that $A_{\bar{w}}\hat{x} > b_{\bar{w}}$, so that $\hat{x}$ is feasible. Do we have a solution to our minimization problem?

Suppose we find that $\hat{\lambda}_j < 0$.

Let p solve $A_w p = e_j$.

(This has a solution since A_w is full rank.)

Then

$$A_w(\hat{x} + \alpha p) = b_w + \alpha e_j \geq b_w,$$

so $\hat{x} + \alpha p$ satisfies the $\mathcal{W}$ **inequality** constraints as long as $\alpha > 0$, and it satisfies the other inequalities as long as α is small enough. Thus, p is a feasible direction.

Also, by Digression 2, we know that

$$f(\hat{x} + \alpha p) \approx f(\hat{x}) + \alpha e_j^T \hat{\lambda} = f(\hat{x}) + \alpha \hat{\lambda}_j < f(\hat{x})$$

(for small enough α) **so we have found a better point!**

We'll come back to the **algorithmic use** of this idea later. For now, we seek insight on recognizing an optimal point.

We have just shown that if x is a minimizer, then the multipliers λ_w that satisfy $A_w^T \lambda_w = g(x)$ must be nonnegative.
(The multipliers for the $\bar{w}$ indices must be zero, since these constraints do not appear in the Lagrangian.)

A fancy way of writing this

Current formulation of (first order) necessary conditions for optimality:

$$\begin{array}{rcl} A_w^T \lambda_w & = & g(x) \\ \lambda_w \geq 0 & , & \lambda_{\bar{w}}=0 \\ A_w x & = & b_w \\ A_{\bar{w}} x & > & b_{\bar{w}} \end{array}$$

where $\bar{w}$ denotes the subscripts not in $\mathcal{W}$.

Equivalently,

$$\begin{aligned} A^T\lambda &= g(x) \\ \lambda &\geq 0 \\ Ax &\geq b \\ \lambda^T(Ax-b) &= 0 \end{aligned}$$

This last condition is called **complementarity**.

The second order necessary condition: (from the reduced variable derivation above) The reduced variable Hessian matrix

$$Z_w^T H(x) Z_w$$

must be positive semidefinite.

Sufficient conditions for optimality: All of this, plus $Z_w^T H(x) Z_w$ positive definite.

Case 3: Optimality conditions for general constraints

$$\min_x f(x)$$

$$c(x) \geq 0$$

A constraint qualification

Let the $m \times n$ matrix $A(x)$ be defined by

$$a_{ij}(x) = \frac{\partial c_i(x)}{\partial x_j}.$$

Assume that $A(x)$ has linearly independent rows.

Again, this is a **constraint qualification**, saying that the gradients of the active constraints are linearly independent.

picture.

Optimality conditions

$$L(x, \lambda) = f(x) - \lambda^T c(x)$$

Theorem: Necessary conditions for a feasible point x to be a minimizer:

- $g(x) - A^T(x)\lambda = 0$
- $\lambda_j \geq 0$ if j is an inequality constraint.
- λ_j unrestricted in sign for equality constraints.
- $\lambda^T c(x) = 0$ (**complementarity**)
- $Z^T \bigtriangledown_{xx} L(x, \lambda) Z$ is positive semidefinite, where the columns of Z are a basis for the null space of A_w, the gradients of the active constraints.

Theorem: Sufficient conditions: Add positive definiteness of $Z^T \bigtriangledown_{xx} L(x, \lambda) Z$.

We won't prove these theorems, but we will sketch the proof of a piece of a special case: that for equality constraints, if x^* is a local minimizer of f, then there is a vector of multipliers satisfying

$$A^T(x^*)\lambda = g(x^*) .$$

Goal:

To prove: If all constraints are equalities, then

$$A^T(x^*)\lambda = g(x^*) .$$

Note: We are proving the correctness of the Lagrange multiplier formulation for solving equality constrained problems as promised earlier in this set of notes.

Proof ingredient 1: a pitfall

With nonlinear constraints, there may be no feasible directions!

picture

So we need to work with **feasible curves** $x(t)$, $0 \le t \le t_1$, with $x(0)$ being our current point. A curve is feasible if it stays tangent to our (active) constraints.

Example 1: The curve

$$x(t) = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$$

stays tangent to the unit circle $x_1^2 + x_2^2 = 1$.

This is true since

$$x(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and

$$x'(t) = \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}$$

which is tangent to the circle. []

Example 2: The curve

$$x(t) = \begin{bmatrix} t \\ 2t \end{bmatrix} + \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

stays tangent to the line

$$x_2 - 2x_1 = 4 \,.$$

[]

Proof ingredient 2: Some unstated machinery that N&S use:

- For $x(t)$ to be a feasible curve, it must be defined for $t \in [t_0, t_1]$, where $t_0 < 0 < t_1$.
- **Every** feasible point in a neighborhood of the current point is on some feasible curve.

Proof ingredient 3: the tangent cone

Define the **tangent cone**

$$T(x^*) = \{p : p = x'(0) \text{for some feasible curve at } x^*\}\,.$$

This is a **cone** because

- $0 \in T$ (because we could define the curve $x(t) = x^*$ for all t).
- If $p \in T$, then $\alpha p \in T$ for positive scalars α.

picture

Now the constraints are equalities, so

$$c_i(x(t)) = 0,\;\; t \in [t_0, t_1]\,,$$

so

$$\frac{d\,c_i(x(t))}{dt} = x'(t)^T \bigtriangledown c_i(x(t)) = 0.\;\; t \in [t_0, t_1]\,.$$

Therefore, at $t = 0$, for all feasible curves,

$$x'(0)^T \bigtriangledown c_i(x^*) = 0\,.$$

Thus, for all p in the tangent cone T of x^*,

$$p^T \bigtriangledown c_i(x^*) = 0,\;\; i = 1, \ldots, m,$$

so

$$A(x^*)p = 0\,.$$

Therefore, if p is in the tangent cone, then p is in the null space of the matrix of constraint gradients!

If the rows of A are linearly independent, then we can reverse the argument and show that **if p is in the null space of A, then p is in the tangent cone.**

Therefore, when $A(x^*)$ is full rank, the tangent cone $T(x^*)$ equals the nullspace of $A(x^*)$.

Finally, the sketch of proof for equality constraints

Suppose x^* is a local minimizer of $f(x)$ over $\{x : c(x){=}0\}$.

Then, for all feasible curves $x(t)$ with $x(0) = x^*$, it must be true that

$$f(x(t)) \geq f(x^*)$$

for $t > 0$ sufficiently small.

The chain rule tells us

$$\frac{d}{dt} f(x(t)) = x'(t)^T \nabla_x f(x(t)),$$

and optimality implies that

$$\frac{d}{dt} f(x(t))|_{t=0} = x'(0)^T \nabla_x f(x^*) = 0.$$

Therefore $p^T g(x^*) = 0$ for all p in the nullspace of $A(x^*)$.

Therefore, a necessary condition for optimality is that the reduced gradient is zero:

$$Z(x^*)^T g(x^*) = 0.$$

Equivalently, there must be a vector λ so that

$$A(x^*)^T \lambda = g(x^*)$$

so that $g(x^*)$ is in the span of the constraint gradients.

[]

picture

Notes on the proof for inequality constraints

- To prove the sign conditions on λ, the argument is the same as for linear constraints.
- To prove the second derivative conditions, see N&S p. 461.

Duality

Duality

Idea: Problems come in pairs, linked through the Lagrangian.

We need two theorems about this linkage, or **duality**:

- weak duality
- strong duality

and then two theorems about **dual problems**:

- weak dual
- convex duality

and finally an alternate dual problem, the Wolfe dual, that depends on differentiability.

Weak duality

Theorem: (Weak Duality) (N&S p466)
Let $F(x, \lambda)$ be a function from $\mathcal{R}^{n+m} \to R^1$ with $x \in \mathcal{R}^n$ and $\lambda \in \mathcal{R}^m$. Then

$$\max_{\lambda} \min_{x} F(x, \lambda) \leq \min_{x} \max_{\lambda} F(x, \lambda) \, .$$

Notes:

- Really, the **max** should be **sup** and the **min** should be **inf**, so substitute this terminology if you are comfortable with it.
- The function F does not need to be defined everywhere; we could restate the theorem with x and λ restricted to smaller domains.

Proof: Given **any** $\hat{x}$ and $\hat{\lambda}$,

$$\min_{x} F(x, \hat{\lambda}) \leq F(\hat{x}, \hat{\lambda}) \leq \max_{\lambda} F(\hat{x}, \lambda) \, .$$

Now let's make a specific choice:

- Let $\hat{\lambda}$ be the λ that maximizes the left-hand side.
- Let $\hat{x}$ be the x that minimizes the right-hand side.

Then

$$\max_{\lambda} \min_{x} F \leq \min_{x} \max_{\lambda} F \, .$$

[]

Strong duality

Theorem: (Strong Duality) (N&S p.468)
Let $F(x, \lambda)$ be a function from $\mathcal{R}^{n+m} \to \mathcal{R}^1$. Then the condition

$$\max_{\lambda} \min_{x} F(x, \lambda) = \min_{x} \max_{\lambda} F(x, \lambda)$$

holds if and only if there exists a point (x^*, λ^*) such that

$$F(x^*, \lambda) \leq F(x^*, \lambda^*) \leq F(x, \lambda^*)$$

for all points x and λ in the domain of F.

In words: We can reverse the order of the max and the min if and only if there exists a saddle point for F.

Proof: $(\leftarrow)$ Suppose (x^*, λ^*) is a saddle point. Then

$$\begin{aligned}
\min_{x} \max_{\lambda} F(x, \lambda) &\leq \max_{\lambda} F(x^*, \lambda) \\
&\leq F(x^*, \lambda^*) \\
&\leq \min_{x} F(x, \lambda^*) \\
&\leq \max_{\lambda} \min_{x} F(x, \lambda)
\end{aligned}$$

Now, considering the result of the weak duality theorem, we can conclude that the first term must **equal** the last.

$(\rightarrow)$ Suppose

$$\max_{\lambda} \min_{x} F(x, \lambda) = \min_{x} \max_{\lambda} F(x, \lambda)$$

and that this is equal to the value $F(x^*, \lambda^*)$. Then, for any $\hat{x}$ and $\hat{\lambda}$,

$$\begin{aligned}
F(x^*, \hat{\lambda}) &\leq \max_{\lambda} F(x^*, \lambda) \\
&= \max_{\lambda} \min_{x} F(x, \lambda) \\
&= F(x^*, \lambda^*) \\
&= \min_{x} \max_{\lambda} F(x, \lambda) \\
&= \min_{x} F(x, \lambda^*) \\
&\leq F(\hat{x}, \lambda^*)
\end{aligned}$$

[]

So what?

Consider our **original problem**:

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

$$c(x) \geq 0$$

The **Lagrangian** for this problem is

$$L(x, \lambda) = f(x) - \lambda^T c(x)\,.$$

A new problem to play with: Lagrange duality

Define

$$L^*(x) = \max_{\lambda \geq 0} L(x, \lambda) = \max_{\lambda \geq 0} f(x) - \lambda^T c(x)\,.$$

Case 1: If x is feasible, then $c(x) \geq 0$, so the **max** occurs when $\lambda = 0$.

Case 2: If x is not feasible, then some $c_i(x)$ is negative, so the max is infinite.

Therefore,

$$L^*(x) = \begin{cases} f(x) & \text{if } c(x) \geq 0\,, \\ \infty & \text{otherwise}\,. \end{cases}$$

Therefore, the solution to the **original problem** is the same as the solution to the **primal problem**

$$\min_x L^*(x) = \min_x \max_{\lambda \geq 0} L(x, \lambda)\,.$$

A dual problem

Suppose $\lambda \geq 0$. Define

$$L_*(\lambda) = \min_x L(x, \lambda) = \min_x f(x) - \lambda^T c(x)\,.$$

Unquiz: Show that the solution to the original problem is the same as the solution to the **dual problem**

$$\max_\lambda L_*(\lambda) = \max_{\lambda \geq 0} \min_x L(x, \lambda)\,.$$

An important example: Linear programming duality

Example: Duality for linear programming

Consider the linear programming problem

$$\min_x c^T x$$

$$Ax - b \geq 0$$

The Lagrangian is

$$L(x, \lambda) = c^T x - \lambda^T (Ax - b)\,.$$

The **primal problem** is

$$\min_x \max_{\lambda \geq 0} c^T x - \lambda^T (Ax - b)$$

which is equivalent to our original problem.

The **dual problem** is

$$\max_{\lambda \geq 0} \min_x c^T x - \lambda^T (Ax - b)\,.$$

Fix $\lambda \geq 0$. Then we need to minimize

$$(c - A^T \lambda)^T x + \lambda^T b$$

and this value is $L_*(\lambda)$.

But

$$L_*(\lambda) = \begin{cases} -\infty & \text{if } c - A^T\lambda \neq 0\,, \\ \lambda^T b & \text{if } c - A^T\lambda = 0\,. \end{cases}$$

Therefore, if $\lambda^* \geq 0$ and $c - A^T\lambda^* = 0$, then the dual problem solution value is $\lambda^{*T} b$.

Thus, the dual problem is equivalent to

$$\max_{\lambda \geq 0} \lambda^T b$$

$$A^T \lambda - c = 0$$

Check strong duality:

Suppose x^* solves the primal and λ^* solves the dual.

Then

$$c^T x^* = \lambda^{*T} b$$

so we can solve either one and know the solution to the other!

For example, if we know λ^*, then the components that are positive determine the **active set** of constraints and enable us to determine x^*.

Remember that the dual variables also give us **sensitivity** information, so they are important to know.

Caution: Usually the variables x and λ **cannot** be uncoupled in the dual. Linear programming is an exception to this.

End of linear programming example. []

Weak Lagrange duality

Theorem: (Weak Lagrange duality) (N&S p. 471)
Let $\tilde{x}$ be primal feasible, so that $c(\tilde{x}) \geq 0$.
Let $\bar{x}, \bar{\lambda}$ be dual feasible, so that $\bar{\lambda} \geq 0$, and $\bar{x}$ minimizes $L(x, \bar{\lambda})$.

Then

$$f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) \leq f(\tilde{x}) .$$

Note:

- For dual feasibility, it is not necessary that $c(x) \geq 0$.
- Sometimes we require that our solution, in addition to satisfying $c(x) \geq 0$, satisfies $x \in S \subset \mathcal{R}^n$. If the problem is formulated this way, then a dual feasible point must have $x \in S$, but it is not necessary that $c(x) \geq 0$.

Proof: Let's recall what we know. The Lagrangian is

$$L(x, \lambda) = f(x) - \lambda^T c(x) .$$

The Weak Duality Theorem, and the fact that $\tilde{x}$ is feasible, tells us

$$\begin{aligned} f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) &= L(\bar{x}, \bar{\lambda}) \\ &\leq \max_{\lambda \geq 0} \min_x L(x, \lambda) \\ &\leq \min_x \max_{\lambda \geq 0} L(x, \lambda) \\ &\leq \max_{\lambda \geq 0} L(\tilde{x}, \lambda) \\ &= f(\tilde{x}) \end{aligned}$$

[]

Corollary: If the primal is unbounded, then the dual is infeasible.
If the dual is unbounded, then the primal is infeasible.

Example: Consider the primal problem

$$\min_x -x$$

(with $x \in \mathcal{R}^1$) subject to $x \geq 0$. The Lagrangian is

$$L(x, \lambda) = -x - \lambda x .$$

Then $\bar{x}, \lambda$ is dual feasible if $\bar{x}$ satisfies

$$\min_x -(\lambda + 1)x$$

where λ is a fixed nonnegative number. There are **no** dual feasible points, and the primal has no minimum. []

Convex Lagrange Duality

Theorem: (Convex duality) (N&S p. 474)
If

- f is convex,
- c_i is concave, $i = 1, \ldots, m$,
- x^* solves the primal,
- and the constraints satisfy a regularity condition at x^*,

then there exists a point λ^* so that x^*, λ^* solves the dual, and the primal and dual function values are equal.

Proof: Let λ^* solve

$$g(x) - A(x)^T \lambda = 0\,.$$

Then

$$\lambda^{*T} c(x^*) = 0\,.$$

1. If x^* is optimal, then $\lambda^* \geq 0$.
2. $L(x, \lambda^*) = f(x) - \lambda^{*T} c(x)$ is convex in x, and x^* minimizes it (since $\nabla_x L = 0$ there), so for all x and λ,

$$f(x^*) = L(x^*, \lambda^*) \leq L(x, \lambda^*),$$

and

$$L(x^*, \lambda^*) \geq L(x^*, \lambda)$$

[]

The Wolfe Dual

If $\bar{x}$ solves

$$\min_x L(x, \lambda)$$

then

$$\nabla_x L(x, \lambda)|_{x=\bar{x}} = 0\,,$$

so we can write the dual as

$$\max_\lambda L(x, \lambda)$$

$$\nabla_x L(x, \lambda) = 0\,.$$

Final words

Final words

- We have derived optimality conditions so that we can recognize a solution when we find one.
- We have derived a partner to our original (primal) problem, called the dual problem.
- We have hinted at some algorithmic approaches:
  - Idea 1: Eliminate constraints by reducing the number of variables.
  - Idea 2: Walk in feasible descent directions.
  - Idea 3: Eliminate constraints through Lagrangians.

Next we will discuss these algorithmic approaches.