Introduction to Convex Constrained Optimization

March 4, 2004

1 Overview

- Nonlinear Optimization
- Portfolio Optimization
- An Inventory Reliability Problem
- Further concepts for nonlinear optimization
- Convex Sets and Convex Functions
- Convex Optimization
- Pattern Classification
- Some Geometry Problems
- On the Geometry of Nonlinear Optimization
- Classification of Nonlinear Optimization Problems
- Solving Separable Convex Optimization via Linear Optimization
- Optimality Conditions for Nonlinear Optimization
- A Few Concluding Remarks

2 Nonlinear versus Linear Optimization

Recall the basic linear optimization model:

$$
\begin{array}{rll}
LP: \quad \text{minimize}_x & c^T x & \\
\text{s.t.} & & \\
& a_1^T x & \le \quad b_1, \\
& \cdot & = \\
& \cdot & \ge \\
& \vdots & \\
& a_m^T x & \le \quad b_m, \\
& x \in \Re^n. &
\end{array}
$$

In this model, all constraints are linear equalities or inequalities, and the objective function is a linear function. In contrast, a nonlinear optimization problem can have nonlinear functions in the constraints and/or the objective function:

$$
\begin{array}{rll}
NLP: \quad \text{minimize}_x & f(x) & \\
\text{s.t.} & & \\
& g_1(x) & \le \quad 0, \\
& \cdot & = \\
& \cdot & \ge \\
& \vdots & \\
& g_m(x) & \le \quad 0, \\
& x \in \Re^n, &
\end{array}
$$

In this model, we have $f(x) : \Re^n \mapsto \Re$ and $g_i(x) : \Re^n \mapsto \Re, i = 1, \ldots, m$.

Below we present several examples of nonlinear optimization models.

3 Portfolio Optimization

Portfolio optimization models are used throughout the financial investment management sector. These are nonlinear models that are used to determine the composition of investment portfolios.

Investors prefer higher annual rates of return on investing to lower annual rates of return. Furthermore, investors prefer lower risk to higher risk. Portfolio optimization seeks to optimally trade off risk and return in investing.

We consider n assets, whose annual rates of return R_i are random variables, $i = 1, \ldots, n$. The expected annual return of asset i is μ_i, $i = 1, \ldots, n$, and so if we invest a fraction x_i of our investment dollar in asset i, the expected return of the portfolio is:

$$\sum_{i=1}^{n} \mu_i x_i = \mu^T x$$

where of course the fractions x_i must satisfy:

$$\sum_{i=1}^{n} x_i = e^T x = 1.0$$

and

$$x \geq 0 \ .$$

(Here, e is the vector of ones, $e = (1, 1, \ldots, 1)^T$.

The covariance of the rates of return of assets i and j is given as

$$Q_{ij} = \mathrm{COV}(i, j) \ .$$

We can think of the Q_{ij} values as forming a matrix Q, whereby the variance of portfolio is then:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{COV}(i, j) x_i x_j = \sum_{i=1}^{n} \sum_{j=1}^{n} Q_{ij} x_i x_j = x^T Q x.$$

It should be noted, by the way, that the matrix Q will always by SPD (symmetric positive-definite).

The "risk" in the portfolio is the standard deviation of the portfolio:

$$\text{STDEV} = \sqrt{x^T Q x}\ ,$$

and the "return" of the portfolio is the expected annual rate of return of the portfolio:

$$\text{RETURN} = \mu^T x\ .$$

Suppose that we would like to determine the fractional investment values $x_1, \ldots, x_n$ in order to maximize the return of the portfolio, subject to meeting some pre-specified target risk level. For example, we might want to ensure that the standard deviation of the portfolio is at most 13.0%. We can formulate the following nonlinear optimization model:

MAXIMIZE:	$\text{RETURN} = \mu^T x$
s.t.	
FSUM:	$e^T x = 1$
ST. DEV.:	$\sqrt{x^T Q x} \quad \leq \quad 13.0$
NONNEGATIVITY:	$x \geq 0$.

An alternative version of the basic portfolio optimization model is to determine the fractional investment values $x_1, \ldots, x_n$ in order to minimize the risk of the portfolio, subject to meeting a pre-specified target expected return level. For example, we might want to ensure that the expected return of the portfolio is at least 16.0%. We can formulate the following nonlinear optimization model:

MINIMIZE:	$\text{STDEV} = \sqrt{x^T Q x}$
s.t.	
FSUM:	$e^t x = 1$
EXP. RETURN:	$\mu^T x \geq\ 16.0$
NONNEGATIVITY:	$x \geq 0$.

Variations of this type of model are used pervasively by asset management companies worldwide. Also, this model can be extended to yield the CAPM (Capital Asset Pricing Model). Finally, we point out the Nobel Prize in economics was awarded in 1990 to Merton Miller, William Sharpe, and Harry Markowitz for their work on portfolio theory and portfolio models (and the implications for asset pricing).

4 An Inventory Reliability Problem

A colleague recently came to me with the follow problem that arises in analyzing inventory management problems:

Given positive coefficients $h_i, \beta_i, d_i, i = 1, \ldots, m$, and $\delta > 0$, solve for $s_1, \ldots, s_m$:

$$
\begin{array}{rlrcll}
IMP: & \text{minimize}_s & \sum\limits_{i=1}^{m} h_i s_i & & & \\
 & \text{s.t.} & & & & \\
 & & \sum\limits_{i=1}^{m} d_i e^{-\beta_i s_i} & \leq & \delta, & \\
 & & s_i & \geq & 0, & i = 1, \ldots, m
\end{array}
$$

5 Further Concepts for Nonlinear Optimization

Recall the basic nonlinear optimization model:

$$
\begin{array}{llrcl}
NLP: & \text{minimize}_x & f(x) & & \\
& \text{s.t.} & & & \\
& & g_1(x) & \leq & 0, \\
& & \cdot & = & \\
& & \cdot & \geq & \\
& & \vdots & & \\
& & g_m(x) & \leq & 0, \\
& & x \in \Re^n, & &
\end{array}
$$

where $f(x) : \Re^n \mapsto \Re$ and $g_i(x) : \Re^n \mapsto \Re, i = 1, \ldots, m$. The *feasible region* of NLP is the set

$$\mathcal{F} = \{x | g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}$$

The *ball* centered at $\bar{x}$ with radius ϵ is the set:

$$B(\bar{x}, \epsilon) := \{x | \|x - \bar{x}\| \leq \epsilon\}$$

We have the following definitions of local/global, strict/non-strict minima/maxima.

Definition 5.1 *$x \in \mathcal{F}$ is a* local minimum *of NLP if there exists $\epsilon > 0$ such that $f(x) \leq f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$.*

Definition 5.2 *$x \in \mathcal{F}$ is a* global minimum *of NLP if $f(x) \leq f(y)$ for all $y \in \mathcal{F}$.*

Definition 5.3 *$x \in \mathcal{F}$ is a* strict local minimum *of NLP if there exists $\epsilon > 0$ such that $f(x) < f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$, $y \neq x$.*

Definition 5.4 *$x \in \mathcal{F}$ is a* strict global minimum *of NLP if $f(x) < f(y)$ for all $y \in \mathcal{F}$, $y \neq x$.*

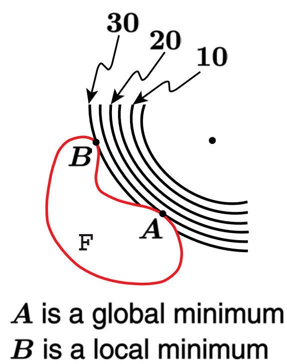

Figure 1: Illustration of local versus global optima.

Definition 5.5 *$x \in \mathcal{F}$ is a* local maximum *of NLP if there exists $\epsilon > 0$ such that $f(x) \geq f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$.*

Definition 5.6 *$x \in \mathcal{F}$ is a* global maximum *of NLP if $f(x) \geq f(y)$ for all $y \in \mathcal{F}$.*

Definition 5.7 *$x \in \mathcal{F}$ is a* strict local maximum *of NLP if there exists $\epsilon > 0$ such that $f(x) > f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$, $y \neq x$.*

Definition 5.8 *$x \in \mathcal{F}$ is a* strict global maximum *of NLP if $f(x) > f(y)$ for all $y \in \mathcal{F}$, $y \neq x$.*

The phenomenon of local versus global optima is illustrated in Figure 1.

5.1 Convex Sets and Functions

Convex sets and convex functions play an extremely important role in the study of optimization models. We start with the definition of a convex set:

Definition 5.9 *A subset $S \subset \Re^n$ is a* convex set *if*

$$x, y \in S \Rightarrow \lambda x + (1 - \lambda) y \in S$$

for any $\lambda \in [0, 1]$.

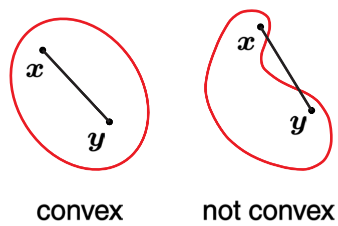

Figure 2: Illustration of convex and non-convex sets.

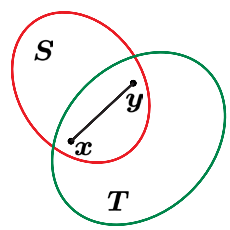

Figure 3: Illustration of the intersection of convex sets.

Figure 2 shows a convex set and a non-convex set.

Proposition 5.1 *If S, T are convex sets, then $S \cap T$ is a convex set.*

This proposition is illustrated in Figure 3.

Proposition 5.2 *The intersection of any collection of convex sets is a convex set.*

We now turn our attention to convex functions, defined below.

Definition 5.10 *A function $f(x)$ is a* convex function *if*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

for all x and y and for all $\lambda \in [0, 1]$.

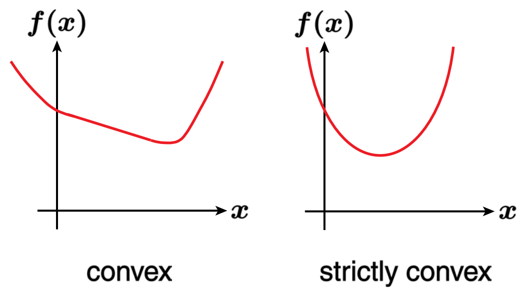

Figure 4: Illustration of convex and strictly convex functions.

Definition 5.11 *A function $f(x)$ is a* strictly convex function *if*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

for all x and y and for all $\lambda \in (0,1)$, $y \neq x$.

Figure 4 illustrates convex and strictly convex functions.

Now consider the following optimization problem, where the feasible region is simply described as the set $\mathcal{F}$:

$$\begin{array}{rll} P: & \text{minimize}_x & f(x) \\ & \text{s.t.} & \\ & & x \in \mathcal{F} \end{array}$$

Proposition 5.3 *Suppose that $\mathcal{F}$ is a convex set, $f : \mathcal{F} \to \Re$ is a convex function, and $\bar{x}$ is a local minimum of P. Then $\bar{x}$ is a global minimum of f over $\mathcal{F}$.*

Proof: Suppose $\bar{x}$ is not a global minimum, i.e., there exists $y \in \mathcal{F}$ for which $f(y) < f(\bar{x})$. Let $y(\lambda) = \lambda\bar{x} + (1-\lambda)y$, which is a convex combination of $\bar{x}$ and y for $\lambda \in [0,1]$ (and therefore, $y(\lambda) \in \mathcal{F}$ for $\lambda \in [0,1]$). Note that $y(\lambda) \to \bar{x}$ as $\lambda \to 1$.

From the convexity of $f(\cdot)$,

$$f(y(\lambda)) = f(\lambda\bar{x}+(1-\lambda)y) \leq \lambda f(\bar{x})+(1-\lambda)f(y) < \lambda f(\bar{x})+(1-\lambda)f(\bar{x}) = f(\bar{x})$$

for all $\lambda \in (0,1)$. Therefore, $f(y(\lambda)) < f(\bar{x})$ for all $\lambda \in (0,1)$, and so $\bar{x}$ is not a local minimum, resulting in a contradiction.
q.e.d.

Some examples of convex functions of one variable are:

- $f(x) = ax + b$
- $f(x) = x^2 + bx + c$
- $f(x) = |x|$
- $f(x) = -\ln(x)$ for $x > 0$
- $f(x) = \frac{1}{x}$ for $x > 0$
- $f(x) = e^x$

5.2 Concave Functions and Maximization

The "opposite" of a convex function is a concave function, defined below:

Definition 5.12 *A function $f(x)$ is a* concave function *if*

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$$

for all x and y and for all $\lambda \in [0,1]$.

Definition 5.13 *A function $f(x)$ is a* strictly concave function *if*

$$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda)f(y)$$

for all x and y and for all $\lambda \in (0,1)$, , $y \neq x$.

Figure 5 illustrates concave and strictly concave functions.

Now consider the maximization problem P:

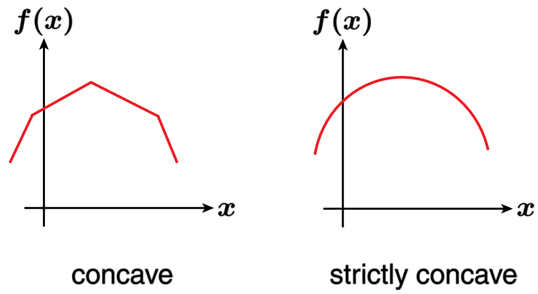

Figure 5: Illustration of concave and strictly concave functions.

$$
\begin{array}{rlll}
P: & \text{maximize}_x & f(x) & \\
& \text{s.t.} & & \\
& & x & \in \mathcal{F}
\end{array}
$$

Proposition 5.4 *Suppose that $\mathcal{F}$ is a convex set, $f : \mathcal{F} \to \Re$ is a concave function, and $\bar{x}$ is a local maximum of P. Then $\bar{x}$ is a global maximum of f over $\mathcal{F}$.*

5.3 Linear Functions, Convexity, and Concavity

Proposition 5.5 *A linear function $f(x) = a^T x + b$ is both convex and concave.*

Proposition 5.6 *If $f(x)$ is both convex and concave, then $f(x)$ is a linear function.*

These properties are illustrated in Figure 6.

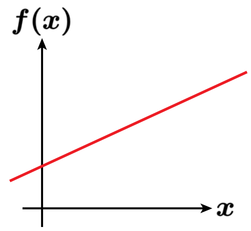

Figure 6: A linear function is convex and concave.

5.4 Convex Optimization

Suppose that $f(x)$ is a convex function. The set

$$S_\alpha := \{x | f(x) \leq \alpha\}$$

is the *level set* of $f(x)$ at level α.

Proposition 5.7 *If $f(x)$ is a convex function, then S_α is a convex set.*

This proposition is illustrated in Figure 7.

Now consider the following optimization:

$$
\begin{array}{rlrcl}
CP: & \text{minimize}_x & f(x) & & \\
& \text{s.t.} & & & \\
& & g_1(x) & \leq & 0, \\
& & \vdots & & \\
& & g_m(x) & \leq & 0, \\
& & Ax & = & b, \\
& & x \in \Re^n, & &
\end{array}
$$

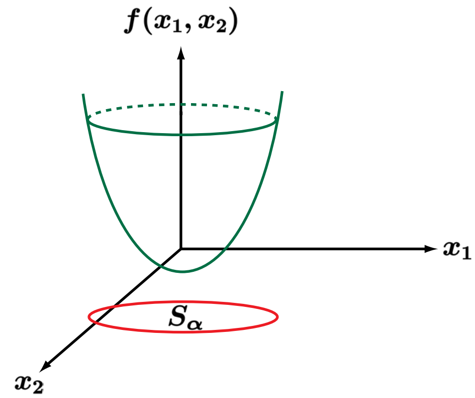

Figure 7: The level sets of a convex function are convex sets.

CP is called a *convex optimization problem* if $f(x), g_1(x), \ldots, g_m(x)$ are convex functions.

Proposition 5.8 *The feasible region of CP is a convex set.*

Proof: From Proposition 5.7, each of the sets

$$\mathcal{F}_i := \{x | g_i(x) \leq 0\}$$

is a convex set, for $i = 1, \ldots, m$. Also, the affine set $\{x | Ax = b\}$ is easily shown to be a convex set. And from Proposition 5.2,

$$\mathcal{F} := \{x | Ax = b\} \cap (\cap_{i=1}^m \mathcal{F}_i)$$

is a convex set.
q.e.d.

Notice therefore that in CP we are minimizing a convex function over a convex set. Applying Proposition 5.3, we have:

Corollary 5.1 *Any local minimum of CP will be a global minimum of CP.*

This is a most important aspect of convex optimization problem.

Remark 1 *If we replace "min" by "max" in CP and if $f(x)$ is a concave function while $g_1(x), \ldots, g_m(x)$ are convex functions, then any local maximum of CP will be a global maximum of CP.*

5.5 Further Properties of Convex Functions

The next two propositions present two very important aspects of convex functions, namely that nonnegative sums of convex functions are convex functions, and that a convex function of an affine transformation of the variables is a convex function.

Proposition 5.9 *If $f_1(x)$ and $f_2(x)$ are convex functions, and $a, b \geq 0$, then*

$$f(x) := af_1(x) + bf_2(x)$$

is a convex function.

Proposition 5.10 *If $f(x)$ is a convex function and $x = Ay + b$, then*

$$g(y) := f(Ay + b)$$

is a convex function.

A function $f(x)$ is *twice differentiable* at $x = \bar{x}$ if there exists a vector $\nabla f(\bar{x})$ (called the *gradient* of $f(\cdot)$) and a symmetric matrix $H(\bar{x})$ (called the *Hessian* of $f(\cdot)$) for which:

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T H(\bar{x})(x - \bar{x}) + R_{\bar{x}}(x)\|x - \bar{x}\|^2$$

where $R_{\bar{x}}(x) \to 0$ as $x \to \bar{x}$.

The gradient vector is the vector of partial derivatives:

$$\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \ldots, \frac{\partial f(\bar{x})}{\partial x_n}\right)^t.$$

The Hessian matrix is the matrix of second partial derivatives:

$$H(\bar{x})_{ij} = \frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j}.$$

The next theorem presents a characterization of a convex function in terms of its Hessian matrix. Recall that *SPSD* means symmetric and positive semi-definite, and *SPD* means symmetric and positive-definite.

Theorem 5.1 *Suppose that $f(x)$ is twice differentiable on the open convex set S. Then $f(x)$ is a convex function on the domain S if and only if $H(x)$ is* SPSD *for all $x \in S$.*

The following functions are examples of convex functions in n-dimensions.

- $f(x) = a^T x + b$
- $f(x) = \frac{1}{2}x^T M x - c^T x$ where M is *SPSD*
- $f(x) = \|x\|$ for any norm $\|\cdot\|$ (see proof below)
- $f(x) = \sum_{i=1}^{m} - \ln(b_i - a_i^T x)$ for x satisfying $Ax < b$.

Corollary 5.2 *If $f(x) = \frac{1}{2}x^T M x - c^T x$ where M is a symmetric matrix, then $f(\cdot)$ is a convex function if and only if M is* SPSD*. Furthermore, $f(\cdot)$ is a strictly convex function if and only if M is* SPD*.*

Proposition 5.11 *The norm function $f(x) = \|x\|$ is a convex function.*

Proof: Let $f(x) := \|x\|$. For any x, y and $\lambda \in [0, 1]$, we have:

$$\begin{aligned} & f(\lambda x + (1-\lambda)y) \\ &= \|\lambda x + (1-\lambda)y\| \\ &\leq \|\lambda x\| + \|(1-\lambda)y\| \\ &= \lambda\|x\| + (1-\lambda)\|y\| \end{aligned}$$

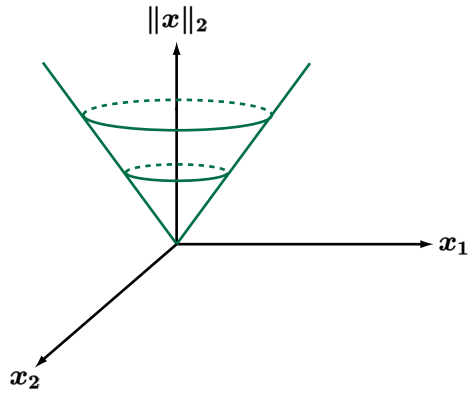

Figure 8: Illustration of a norm function.

$$= \lambda f(x) + (1-\lambda) f(y).$$

q.e.d.

This proposition is illustrated in Figure 8.

When a convex function is differentiable, it must satisfy the following property, called the "gradient inequality".

Proposition 5.12 *If $f(\cdot)$ is a differentiable convex function, then for any x, y,*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \ .$$

Even when a convex function is not differentiable, it must satisfy the following "subgradient" inequality.

Proposition 5.13 *If $f(\cdot)$ is a convex function, then for every x, there must exist some vector s for which*

$$f(y) \geq f(x) + s^T (y - x) \quad \text{for any } y \ .$$

The vector s in this proposition is called a *subgradient* of $f(\cdot)$ at the point x.

6 More Examples of Convex Optimization Problems

6.1 A Pattern Classification Training Problem

We are given:

- points $a^1, \ldots, a^k \in \Re^n$ that have property "P"
- points $b^1, \ldots, b^m \in \Re^n$ that do not have property "P"

We would like to use these $k+m$ points to develop a linear rule that can be used to predict whether or not other points x might or might not have property P. In particular, we seek a vector v and a scalar β for which:

- $v^T a^i > \beta$ for all $i = 1, \ldots, k$
- $v^T b^i < \beta$ for all $i = 1, \ldots, m$

We will then use v, β to predict whether or not any other point c has property P or not. If we are given another vector c, we will declare whether c has property P or not as follows:

- If $v^T c > \beta$, then we declare that c has property P.
- If $v^T c < \beta$, then we declare that c does not have property P.

We therefore seek v, β that define the hyperplane

$$H_{v,\beta} := \{x | v^T x = \beta\}$$

for which:

- $v^T a^i > \beta$ for all $i = 1, \ldots, k$
- $v^T b^i < \beta$ for all $i = 1, \ldots, m$

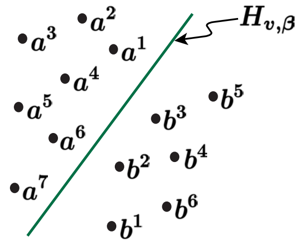

Figure 9: Illustration of the pattern classification problem.

This is illustrated in Figure 9.

In addition, we would like the hyperplane $H_{v,\beta}$ to be as far away from the points $a^1, \ldots, a^k, b^1, \ldots, b^k$ as possible. Elementary analysis shows that the distance of the hyperplane $H_{v,\beta}$ to any point a^i is equal to

$$\frac{v^T a^i - \beta}{\|v\|}$$

and similarly the distance of the hyperplane $H_{v,\beta}$ to any point b^i is equal to

$$\frac{\beta - v^T b^i}{\|v\|}.$$

If we normalize the vector v so that $\|v\| = 1$, then the minimum distance of the hyperplane $H_{v,\beta}$ to the points $a^1, \ldots, a^k, b^1, \ldots, b^k$ is then:

$$\min\{v^T a^1 - \beta, \ldots, v^T a^k - \beta, \beta - v^T b^1, \ldots, \beta - v^T b^m\}$$

Therefore we would also like v and β to satisfy:

- $\|v\| = 1$, and
- $\min\{v^T a^1 - \beta, \ldots, v^T a^k - \beta, \beta - v^T b^1, \ldots, \beta - v^T b^m\}$ is maximized.

This results in the following optimization problem:

$$
\begin{aligned}
PCP: \quad \text{maximize}_{v,\beta,\delta} \quad & \delta \\
\text{s.t.} \quad & \\
& v^T a^i - \beta \geq \delta, \quad i = 1, \ldots, k \\
& \beta - v^T b^i \geq \delta, \quad i = 1, \ldots, m \\
& \|v\| = 1, \\
& v \in \Re^n, \beta \in \Re
\end{aligned}
$$

Now notice that as written, PCP is not a convex optimization problem, due to the presence of the constraint "$\|v\| = 1$." However, if we perform the following transformation of variables:

$$x = \frac{v}{\delta} \quad , \quad \alpha = \frac{\beta}{\delta}$$

then maximizing δ is the same as maximizing $\frac{\|v\|}{\|x\|} = \frac{1}{\|x\|}$, which is the same as minimizing $\|x\|$. Therefore we can write the equivalent problem:

$$
\begin{aligned}
CPCP: \quad \text{minimize}_{x,\alpha} \quad & \|x\| \\
\text{s.t.} \quad & \\
& x^T a^i - \alpha \geq 1, \quad i = 1, \ldots, k \\
& \alpha - x^T b^i \geq 1, \quad i = 1, \ldots, m \\
& x \in \Re^n, \alpha \in \Re
\end{aligned}
$$

Here we see that $CPCP$ is a convex problem. We can solve $CPCP$ for x, α, and substitute $v = \frac{x}{\|x\|}$, $\beta = \frac{\alpha}{\|x\|}$ and $\delta = \frac{1}{\|x\|}$ to obtain the solution of PRP.

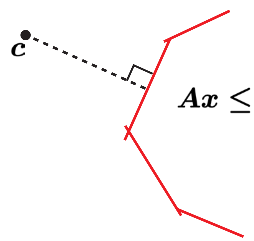

Figure 10: Illustration of the minimum norm problem.

6.2 The Minimum Norm Problem

Given a vector c, we would like to find the closest point to c that also satisfies the linear inequalities $Ax \leq b$.

This problem is:

$$\begin{array}{rll} MNP: & \text{minimize}_x & \|x - c\| \\ & \text{s.t.} & \\ & & Ax \quad \leq \quad b \\ & & x \in \Re^n \end{array}$$

This problem is illustrated in Figure 10.

6.3 The Fermat-Weber Problem

We are given m points $c^1, \ldots, c^m \in \Re^n$. We would like to determine the location of a distribution center at the point $x \in \Re^n$ that minimizes the sum of the distances from x to each of the points $c^1, \ldots, c^m \in \Re^n$. This problem is illustrated in Figure 11. It has the following formulation:

$$
\begin{array}{lll}
FWP: & \text{minimize}_x & \sum_{i=1}^{m} \|x - c^i\| \\
 & \text{s.t.} & \\
 & & x \in \Re^n
\end{array}
$$

Notice that FWP is a convex unconstrained problem.

6.4 The Ball Circumscription Problem

We are given m points $c^1, \ldots, c^m \in \Re^n$. We would like to determine the location of a distribution center at the point $x \in \Re^n$ that minimizes the maximum distance from x to any of the points $c^1, \ldots, c^m \in \Re^n$. This problem is illustrated in Figure 12. It has the following formulation:

$$
\begin{array}{llllll}
BCP: & \text{minimize}_{x,\delta} & \delta & & & \\
 & \text{s.t.} & & & & \\
 & & \|x - c^i\| & \leq & \delta, & i = 1, \ldots, m, \\
 & & x \in \Re^n & & &
\end{array}
$$

This is a convex (constrained) optimization problem.

6.5 The Analytic Center Problem

Given a system of linear inequalities $Ax \leq b$, we would like to determine a "nicely" interior point $\hat{x}$ that satisfies $A\hat{x} < b$. Of course, we would like the point to be as interior as possible, in some mathematically meaningful way. As it turns out, the solution of the following problem has some remarkable properties:

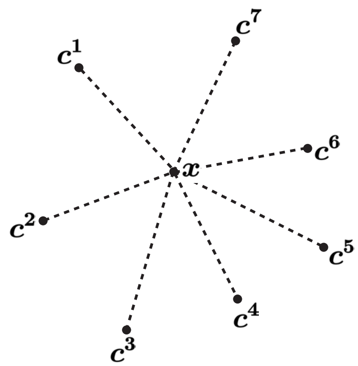

Figure 11: Illustration of the Fermat-Weber problem.

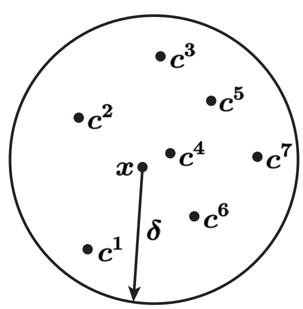

Figure 12: Illustration of the ball circumscription problem.

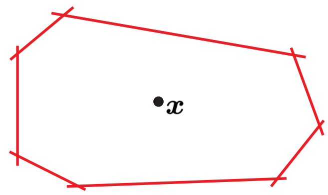

Figure 13: Illustration of the analytic center problem.

$$
\begin{array}{llrcl}
ACP: & \text{maximize}_x & \prod_{i=1}^{m} (b - Ax)_i & & \\
\text{s.t.} & & & & \\
 & & Ax & \leq & b, \\
 & & x \in \Re^n & &
\end{array}
$$

This problem is illustrated in Figure 13.

Proposition 6.1 *Suppose that $\hat{x}$ solves the analytic center problem ACP. Then for each i, we have:*

$$(b - A\hat{x})_i \geq \frac{s_i^*}{m}$$

where

$$s_i^* := \max_{Ax \leq b} (b - Ax)_i \ .$$

Notice that as stated, ACP is not a convex problem, but it is equivalent to:

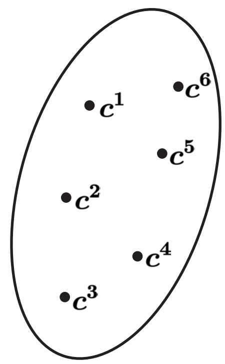

Figure 14: Illustration of the circumscribed ellipsoid problem.

$$
\begin{array}{llll}
CACP: & \text{minimize}_x & \sum_{i=1}^{m} -\ln((b - Ax)_i) & \\
& \text{s.t.} & & \\
& & Ax & < \; b, \\
& & x \in \Re^n &
\end{array}
$$

Now notice that $CACP$ is a convex problem.

6.6 The Circumscribed Ellipsoid Problem

We are given m points $c^1, \ldots, c^m \in \Re^n$. We would like to determine an ellipsoid of minimum volume that contains each of the points $c^1, \ldots, c^m \in \Re^n$. This problem is illustrated in Figure 14.

Before we show the formulation of this problem, first recall that an *SPD* matrix R and a given point z can be used to define an ellipsoid in $\Re^n$:

$$E_{R,z} := \{y \mid (y - z)^T R(y - z) \leq 1\}.$$

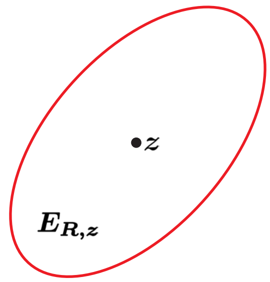

Figure 15: Illustration of an ellipsoid.

Figure 15 shows an illustration of an ellipsoid.

One can prove that the volume of $E_{R,z}$ is proportional to $\det(R^{-1})^{\frac{1}{2}}$.

Our problem is:

$$
\begin{array}{lll}
MCP_1: & \underset{R,z}{\text{minimize}} & \det(R^{-1})^{\frac{1}{2}} \\
& \text{s.t.} & c_i \in E_{R,z}, \quad i = 1, \ldots, k, \\
& & \\
& & R \text{ is SPD.}
\end{array}
$$

Now minimizing $\det(R^{-1})^{\frac{1}{2}}$ is the same as minimizing $\ln(\det(R^{-1})^{\frac{1}{2}})$, since the logarithm function is strictly increasing in its argument. Also,

$$\ln(\det(R^{-1})^{\frac{1}{2}}) = -\frac{1}{2}\ln(\det(R))$$

and so our problem is equivalent to:

$$
\begin{array}{lll}
MCP_2: & \underset{R,z}{\text{minimize}} & -\ln(\det(R)) \\
& \text{s.t.} & (c_i - z)^T R (c_i - z) \le 1, \quad i = 1, \ldots, k \\
& & \\
& & R \text{ is SPD .}
\end{array}
$$

We now factor $R = M^2$ where M is SPD (that is, M is a square root of R), and now MCP becomes:

$$\begin{array}{lll} MCP_3: & \underset{M,z}{\text{minimize}} & -\ln(\det(M^2)) \\ & \text{s.t.} & (c_i - z)^T M^T M (c_i - z) \leq 1, \quad i = 1, \dots, k, \\ & & \\ & & M \text{ is SPD.} \end{array}$$

which is the same as:

$$\begin{array}{lll} MCP_4: & \underset{M,z}{\text{minimize}} & -2\ln(\det(M)) \\ & \text{s.t.} & \|M(c_i - z)\| \leq 1, \quad i = 1, \dots, k, \\ & & \\ & & M \text{ is SPD.} \end{array}$$

Next substitute $y = Mz$ to obtain:

$$\begin{array}{lll} MCP_5: & \underset{M,y}{\text{minimize}} & -2\ln(\det(M)) \\ & \text{s.t.} & \|Mc_i - y\| \leq 1, \quad i = 1, \dots, k, \\ & & \\ & & M \text{ is SPD.} \end{array}$$

It turns out that this is convex problem in the variables M, y.

We can recover R and z after solving MCP_5 by substituting $R = M^2$ and $z = M^{-1}y$.

7 Classification of Nonlinear Optimization Problems

7.1 General Nonlinear Problem

$$\begin{array}{ll} \text{min} \quad \text{or} \quad \text{max} & f(x) \\ \text{s.t.} & g_i(x) \leq b_i \ , \quad i = 1, \dots, m \end{array}$$

7.2 Unconstrained Optimization

$$\min \quad \text{or} \quad \max \quad f(x)$$

$$\text{s.t.} \quad x \in \Re^n$$

7.3 Linearly Constrained Problems

$$\min \quad \text{or} \quad \max \quad f(x)$$

$$\text{s.t.} \quad Ax \leq b$$

7.4 Linear Optimization

$$\min \quad c^T x$$

$$\text{s.t.} \quad Ax \leq b$$

7.5 Quadratic Problem

$$\min \quad c^T x + \tfrac{1}{2} x^T Q x$$

$$\text{s.t.} \quad Ax \leq b$$

The objective function here is $\min \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} \sum_{k=1}^{n} x_j Q_{jk} x_k$.

7.6 Quadratically Constrained Quadratic Problem

$$\min \quad c^T x + \tfrac{1}{2} x^T Q x$$

$$\text{s.t.} \quad a_i^T x + \tfrac{1}{2} x^T Q_i x \leq b_i \ , \quad i = 1, \ldots, m$$

7.7 Separable Problem

$$\min \quad \sum_{j=1}^{n} f_j(x)$$

$$\text{s.t.} \quad \sum_{j=1}^{n} g_{ij}(x_j) \leq b_i \ , \quad i = 1, \ldots, m$$

7.7.1 Example of a separable problem

$$\begin{aligned} \min \quad & x^2 + 2^y + \sqrt{1+z^3} \\ \text{s.t.} \quad & x^2 + y^2 + \cos z \leq 9 \\ & \sin x + \sqrt{3-y} + \sin(z+\pi) \leq 18 \\ & x \geq 0, y \leq 0, z \geq 0 \end{aligned}$$

7.8 Convex Problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq b_i \ , \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where $f(x)$ is a convex function and $g_1(x), \ldots, g_m(x)$ are convex functions, or:

$$\begin{aligned} \max \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq b_i \ , \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where $f(x)$ is a concave function and $g_1(x), \ldots, g_m(x)$ are convex functions.

7.8.1 Example of a convex problem

$$\begin{aligned} \min \quad & x^2 + e^y \\ \text{s.t.} \quad & x^2 + y^2 \le 64 \\ & x + y \le 9 \\ & (x-10)^2 + (y)^2 \le 25 \\ & x \ge 0, y \ge 0 \end{aligned}$$

8 Solving Separable Convex Optimization via Linear Optimization

Consider the problem

$$\begin{aligned} \min \quad & (x_1 - 2)^2 + 2^{x_2} + 9x_3 \\ \text{s.t.} \quad & 3x_1 + 2x_2 + 5x_3 \le 7 \\ & 7x_1 + 5x_2 + x_3 \le 8 \\ & x_1, x_2, x_3 \ge 0 \end{aligned}$$

This problem has:

1. linear constraints
2. a separable objective function $z = f_1(x_1) + f_2(x_2) + 9x_3$
3. a convex objective function (since $f_1(\cdot)$ and $f_2(\cdot)$ are convex functions)

In this case, we can approximate $f_1(\cdot)$ and $f_2(\cdot)$ by piecewise-linear (PL) functions. Suppose we know that $0 \le x_1 \le 9$ and $0 \le x_2 \le 9$. Referring to Figure 16 and Figure 17, we see that we can approximate these functions as follows:

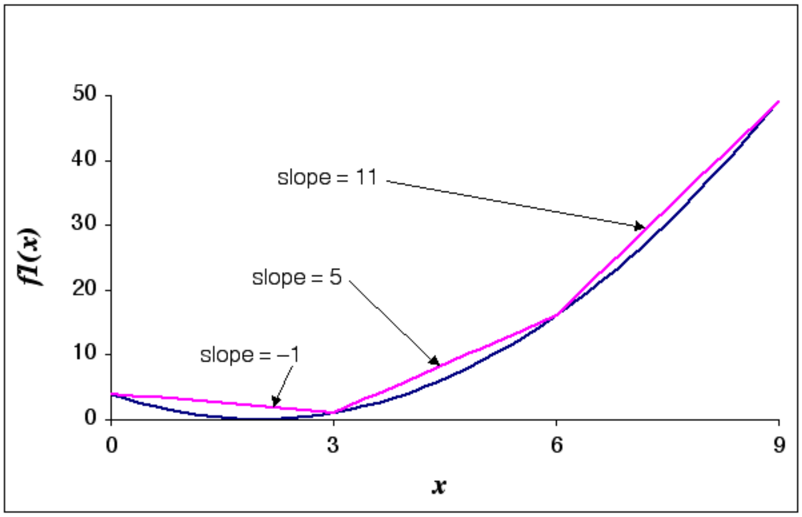

Figure 16: PL approximation of $f_1(\cdot)$.

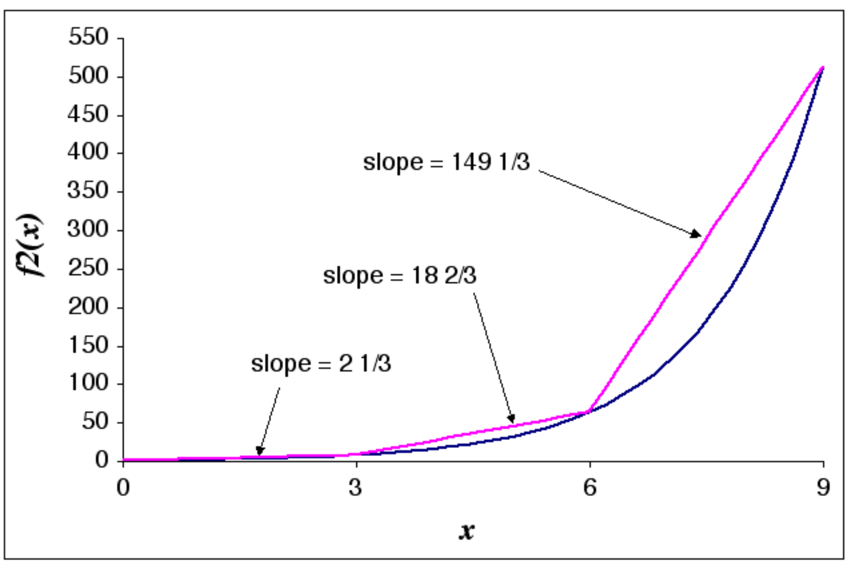

Figure 17: PL approximation of $f_2(\cdot)$.

$$f_1(x_1) = (x_1 - 2)^2 \approx 4 - x_{1a} + 5x_{1b} + 11x_{1c}$$

where

$$x_1 = x_{1a} + x_{1b} + x_{1c}$$

and:

$$0 \le x_{1a} \le 3, \quad 0 \le x_{1b} \le 3, \quad 0 \le x_{1c} \le 3 \ .$$

Similarly,

$$f_2(x_2) = 2^{x_2} \approx 1 + \frac{7}{3}x_{2a} + 18\frac{2}{3}x_{2b} + 149\frac{1}{3}x_{2c}$$

where

$$x_2 = x_{2a} + x_{2b} + x_{2c}$$

and:

$$0 \le x_{2a} \le 3, \quad 0 \le x_{2b} \le 3, \quad 0 \le x_{2c} \le 3$$

The linear optimization approximation then is:

$$
\begin{aligned}
\min \quad & 4 - x_{1a} + 5x_{1b} + 11x_{1c} + 1 + \tfrac{7}{3}x_{2a} + 18\tfrac{2}{3}x_{2b} + 149\tfrac{1}{3}x_{2c} + 9x_3 \\
\text{s.t.} \quad & 3x_1 + 2x_2 + 5x_3 \leq 7 \\
& 7x_1 + 5x_2 + x_3 \leq 8 \\
& x_1, x_2, x_3 \geq 0 \\
& x_{1a} + x_{1b} + x_{1c} = x_1 \\
& x_{2a} + x_{2b} + x_{2c} = x_2 \\
& 0 \leq x_{1a} \leq 3 \\
& 0 \leq x_{1b} \leq 3 \\
& 0 \leq x_{1c} \leq 3 \\
& 0 \leq x_{2a} \leq 3 \\
& 0 \leq x_{2b} \leq 3 \\
& 0 \leq x_{2c} \leq 3
\end{aligned}
$$

9 On the Geometry of Nonlinear Optimization

Figures 18, 19, and 20 show three possible configurations of optimal solutions of nonlinear optimization models. These figures illustrate that unlike linear optimization, the optimal solution of a nonlinear optimization problem need not be a "corner point" of $\mathcal{F}$. The optimal solution may be on the boundary or even in the interior of $\mathcal{F}$. However, as all three figures show, the optimal solution will be characterized by how the contours of $f(\cdot)$ are "aligned" with the feasible region $\mathcal{F}$.

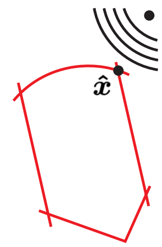

Figure 18: First example of the geometry of the solution of a nonlinear optimization problem.

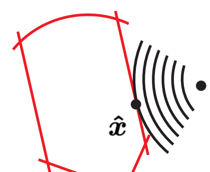

Figure 19: Second example of the geometry of the solution of a nonlinear optimization problem.

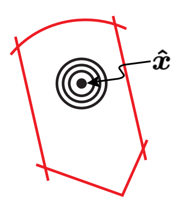

Figure 20: Third example of the geometry of the solution of a nonlinear optimization problem.

10 Optimality Conditions for Nonlinear Optimization

Consider the convex problem:

$$\begin{array}{lll} (\mathrm{CP}): & \min & f(x) \\ \\ \text{s.t.} & & g_i(x) \le b_i \ , \quad i = 1, \ldots, m \ . \end{array}$$

where $f(x), g_i(x)$ are convex functions.

We have:

Theorem 10.1 *(Karush-Kuhn-Tucker Theorem) Suppose that $f(x), g_1(x), \ldots, g_m(x)$ are all convex functions. Then under very mild conditions, $\bar{x}$ solves (CP) if and only if there exists $\bar{y}_i \ge 0, i = 1, \ldots, m$, such that*

$$\begin{array}{lll} (i) & \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{y}_i \nabla g_i(\bar{x}) = 0 & \textit{(gradients line up)} \\ \\ (ii) & g_i(\bar{x}) - b_i \le 0 & \textit{(feasibility)} \\ \\ (iii) & \bar{y}_i \ge 0 & \textit{(multipliers must be nonnegative)} \\ \\ (iv) & \bar{y}_i(b_i - g_i(\bar{x})) = 0. & \textit{(multipliers for nonbinding constraints are zero)} \end{array}$$

10.1 The KKT Theorem generalizes linear optimization strong duality

Let us see how the KKT Theorem applies to linear optimization. Consider the following linear optimization problem and its dual:

$$\begin{array}{lllcll} (\mathrm{LP}): & \min & c^T x & \quad & (\mathrm{LD}): \ \max & \sum_{i=1}^{m} -y_i b_i \\ \\ & & a_i^T x \le b_i \ , \ i = 1, \ldots, m & & & -\sum_{i=1}^{m} y_i a_i = c \\ \\ & & & & & y \ge 0 \end{array}$$

Now let us look at the KKT Theorem applied to this problem. The KKT Theorem states that if $\bar{x}$ solves LP, then there exists $\bar{y}_i, i = 1, \ldots, m$ for which:

$$
\begin{array}{lll}
(i) & \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{y}_i \nabla g_i(\bar{x}) := c + \sum_{i=1}^{m} \bar{y}_i a_i = 0 & \text{(gradients line up)} \\
(ii) & g_i(\bar{x}) - b_i := a_i^T x - b_i \leq 0 & \text{(primal feasibility)} \\
(iii) & \bar{y}_i \geq 0 & \\
(iv) & \bar{y}_i(b_i - g_i(\bar{x})) := \bar{y}_i(b_i - a_i^T x) = 0. & \text{(complementary slackness)}
\end{array}
$$

Now notice that (ii) is primal feasibility of $\bar{x}$, and (i) and (iii) together are dual feasibility of $\bar{y}$. Finally, (iv) is complementary slackness. Therefore, the KKT Theorem here states that $\bar{x}$ and $\bar{y}$ together must be primal feasible and dual feasible, and must satisfy complementary slackness.

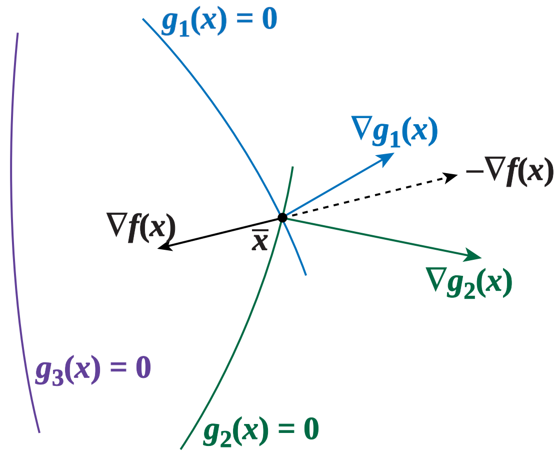

Figure 21: Illustration of the KKT Theorem.

10.2 Geometry of the KKT Theorem

10.3 An example of the KKT Theorem

Consider the problem:

$$
\begin{array}{llll}
\min & 6(x_1-10)^2 & +4(x_2-12.5)^2 & \\
\text{s.t.} & x_1^2 & +(x_2-5)^2 & \leq 50 \\
 & x_1^2 & +3x_2^2 & \leq 200 \\
 & (x_1-6)^2 & +x_2^2 & \leq 37
\end{array}
$$

In this problem, we have:

$$f(x) = 6(x_1-10)^2 + 4(x_2-12.5)^2$$

$$g_1(x) = x_1^2 + (x_2-5)^2$$

$$g_2(x) = x_1^2 + 3x_2^2$$

$$g_3(x) = (x_1-6)^2 + x_2^2$$

We also have:

$$\nabla f(x) = \begin{pmatrix} 12(x_1-10) \\ \\ 8(x_2-12.5) \end{pmatrix}$$

$$\nabla g_1(x) = \begin{pmatrix} 2x_1 \\ \\ 2(x_2-5) \end{pmatrix}$$

$$\nabla g_2(x) = \begin{pmatrix} 2x_1 \\ \\ 6x_2 \end{pmatrix}$$

$$\nabla g_3(x) = \begin{pmatrix} 2(x_1 - 6) \\ \\ 2x_2 \end{pmatrix}$$

Let us determine whether or not the point $\bar{x} = (\bar{x}_1, \bar{x}_2) = (7, 6)$ is an optimal solution to this problem.

We first check for feasibility:

$$g_1(\bar{x}) = 50 \leq 50 = b_1$$

$$g_2(\bar{x}) = 157 < 200 = b_2$$

$$g_3(\bar{x}) = 37 \leq 37 = b_3$$

To check for optimality, we compute all gradients at $\bar{x}$:

$$\nabla f(x) = \begin{pmatrix} -36 \\ \\ -52 \end{pmatrix}$$

$$\nabla g_1(x) = \begin{pmatrix} 14 \\ \\ 2 \end{pmatrix}$$

$$\nabla g_2(x) = \begin{pmatrix} 14 \\ \\ 36 \end{pmatrix}$$

$$\nabla g_3(x) = \begin{pmatrix} 2 \\ \\ 12 \end{pmatrix}$$

We next check to see if the gradients "line up", by trying to solve for $y_1 \geq 0, y_2 = 0, y_3 \geq 0$ in the following system:

$$\begin{pmatrix} -36 \\ \\ -52 \end{pmatrix} + \begin{pmatrix} 14 \\ \\ 2 \end{pmatrix} y_1 + \begin{pmatrix} 14 \\ \\ 36 \end{pmatrix} y_2 + \begin{pmatrix} 2 \\ \\ 12 \end{pmatrix} y_3 = \begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}$$

Notice that $\bar{y} = (\bar{y}_1, \bar{y}_2, \bar{y}_3) = (2, 0, 4)$ solves this system in nonnegative values, and that $y_2 = 0$. Therefore $\bar{x}$ is an optimal solution to this problem.

10.4 The KKT Theorem with Different formats of Constraints

Suppose that our optimization problem is of the following form:

$$\begin{array}{lll} (\text{CPE}): & \min & f(x) \\ \text{s.t.} & & g_i(x) \le b_i\ , \quad i \in I \\ & & a_i^T x = b_i\ , \quad i \in E \end{array}$$

The KKT Theorem for this model is as follows:

Theorem 10.2 *(Karush-Kuhn-Tucker Theorem) Suppose that $f(x), g_i(x)$ are all convex functions for $i \in I$. Then under very mild conditions, $\bar{x}$ solves (CPE) if and only if there exists $\bar{y}_i \ge 0, i \in I$ and $\bar{v}_i, i \in E$ such that*

$$\begin{array}{lll} (i) & \nabla f(\bar{x}) + \sum\limits_{i \in I} \bar{y}_i \nabla g_i(\bar{x}) + \sum\limits_{i \in E}^{k} \bar{v}_i a_i = 0 & \textit{(gradients line up)} \\ (ii-a) & g_i(\bar{x}) - b_i \le 0,\ i \in I & \textit{(feasibility)} \\ (ii-b) & a_i^t x - b_i = 0,\ i \in E & \textit{(feasibility)} \\ (iii) & \bar{y}_i \ge 0,\ i \in I & \textit{(nonnegative multipliers on inequalities)} \\ (iv) & \bar{y}_i(b_i - g_i(\bar{x})) = 0,\ i \in I. & \textit{(complementary slackness)} \end{array}$$

What about the non-convex case? Let us consider a problem of the following form:

$$\begin{array}{llr}(\text{NCE}): & \min & f(x) \\ \text{s.t.} & & g_i(x) \leq b_i \ , i \in I \\ & & h_i(x) = b_i \ , i \in E\end{array}$$

Well, when the problem is not convex, we can at least assert that any optimal solution must satisfy the KKT conditions:

Theorem 10.3 *(Karush-Kuhn-Tucker Theorem) Under some very mild conditions, if $\bar{x}$ solves (NCE), then there exists $\bar{y}_i \geq 0, i \in I$ and $\bar{v}_i, i \in E$ such that*

$$\begin{array}{lll}(i) & \nabla f(\bar{x}) + \sum\limits_{i \in I} \bar{y}_i \nabla g_i(\bar{x}) + \sum\limits_{i \in E} \bar{v}_i \nabla h_i(x) = 0 & \textit{(gradients line up)} \\ (ii-a) & g_i(\bar{x}) - b_i \leq 0, \ i \in I & \textit{(feasibility)} \\ (ii-b) & h_i(x) - b_i = 0, \ i \in E & \textit{(feasibility)} \\ (iii) & \bar{y}_i \geq 0 & \textit{(nonnegative multipliers on inequalities)} \\ (iv) & \bar{y}_i(b_i - g_i(\bar{x})) = 0. & \textit{(complementary slackness)}\end{array}$$

11 A Few Concluding Remarks

1. Nonconvex optimization problems can be difficult to solve. This is because local optima may not be global optima. Most algorithms are based on calculus, and so can only find a local optimum.

2. Solution Methods. There are a large number of solution methods for solving nonlinear constrained problems. Today, the most powerful methods fall into two classes:

   - methods that try to generalize the simplex algorithm to the nonlinear case.
   - methods that generalize the barrier method to the nonlinear case.

3. Quadratic Problems. A quadratic problem is “almost linear”, and can be solved by a special implementation of the simplex algorithm, called the complementary pivoting algorithm. Quadratic problems are roughly eight times harder to solve than linear optimization problems of comparable size.