Lecture 3

1. Klee-Minty Example
2. Convexity
3. Analytic center
4. Central path

Klee-Minty Example

One form of the Klee–Minty example is

$$
\begin{aligned}
&\text{maximize} && \sum_{j=1}^{n} 10^{n-j} x_j \\
&\text{subject to} && 2\sum_{j=1}^{i-1} 10^{i-j} x_j + x_i \le 100^{i-1} \quad i = 1, \ldots, n \\
& && x_j \ge 0 \qquad j = 1, \ldots, n
\end{aligned}
\tag{1}
$$

The problem above is easily cast as a linear program in standard form.
A specific case is that for $n = 3$, giving

$$
\begin{array}{lrcrcrcr}
\text{maximize} & 100x_1 & + & 10x_2 & + & x_3 & & \\
\text{subject to} & x_1 & & & & & \le & 1 \\
& 20x_1 & + & x_2 & & & \le & 100 \\
& 200x_1 & + & 20x_2 & + & x_3 & \le & 10{,}000 \\
& x_1 \geqslant 0, & & x_2 \geqslant 0, & & x_3 \geqslant 0. & &
\end{array}
$$

Tableaux for Klee Minty

- Let us check the corresponding tableaux.

1	0	0	1	0	0	1
20	1	0	0	1	0	100
200	20	1	0	0	1	10000
100	10	1	0	0	0	0

1	0	0	1	0	0	1
0	1	0	-20	1	0	80
0	20	1	-200	0	1	9800
0	10	1	-100	0	0	-100

1	0	0	1	0	0	1
0	1	0	-20	1	0	80
0	0	1	200	-20	1	8200
0	0	1	100	-10	0	-900

1	0	0	1	0	0	1
20	1	0	0	1	0	100
-200	0	1	0	-20	1	8000
-100	0	1	0	-10	0	-1000

Tableaux for Klee Minty

1	0	0	1	0	0	1
20	1	0	0	1	0	100
−200	0	1	0	−20	1	8000
100	0	0	0	10	−1	−9000

1	0	0	1	0	0	1
0	1	0	−20	1	0	80
0	0	1	200	−20	1	8200
0	0	0	−100	10	−1	−9100

1	0	0	1	0	0	1
0	1	0	−20	1	0	80
0	20	1	−200	0	1	9800
0	−10	0	100	0	−1	−9900

1	0	0	1	0	0	1
20	1	0	0	1	0	100
200	20	1	0	0	1	10000
−100	−10	0	0	0	−1	−10000

- Suppose now that we do a simple change of variables:

$$u_j = 100^{1-j} x_j$$

- that is $u_1 = x_1, 100u_2 = x_2$ and $10000u_3 = x_3$
- This is just a scaling of the variables and should not (ideally) affect the complexity of the problem
- The constraint become:

$$\begin{array}{rcrcrcr} u_1 & & & & & \leq & 1 \\ 20u_1 & + & 100u_2 & & & \leq & 100 \\ 200u_1 & + & 2000u_2 & + & 10000u_3 & \leq & 10000. \end{array}$$

- The objective: maximize $100u_1 + 1000u_2 + 10000u_3$.

- Only one pivot

1	0	0	1	0	0	1
20	100	0	0	1	0	100
200	2000	10000	0	0	1	10000
100	1000	10000	0	0	0	0

1	0	0	1	0	0	1
20	100	0	0	1	0	100
0	0	1	0	0	0	1
−100	−1000	0	0	0	−1	−10000

- After the change of variable, the simplex method performs much better...
- This means that the **largest reduced cost coefficient** rule is probably not the most reasonable choice.

Simplex: efficiency

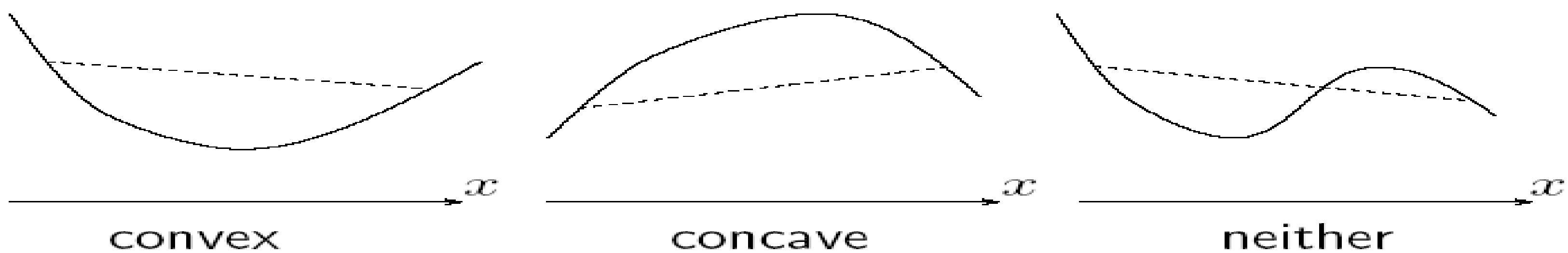
- Klee and Minty show that the largest coefficient rule takes $2^m - 1$ pivots to solve a given problem with m variables and constraints.
- For $m = 70$, this means
$$2^m = 1.2\ 10^{21} \text{ pivots}$$
- At 1000 iterations per second, it will take 40 billion years to solve the problem. (The age of the universe is estimated at 15 billion years)
- On the other hand, very large problems are solved routinely with $m = 10,000$.
- Conclusion here: simplex can take an exponential amount of time on pathological problems.

Convex functions

$f : \mathbf{R}^n \rightarrow \mathbf{R}$ is **convex** if $\mathbf{dom}\, f$ is convex and for all $x, y \in \mathbf{dom}\, f$, $\theta \in [0,1]$

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

f is concave if $-f$ is convex



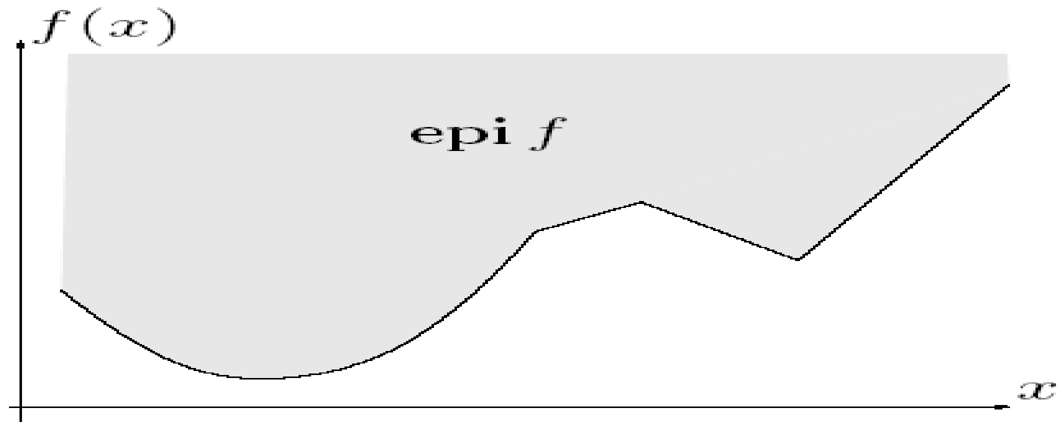
examples (on $\mathbf{R}$)

- $f(x) = x^2$ is convex
- $f(x) = \log x$ is concave ($\mathbf{dom}\, f = \{x | x > 0\}$)
- $f(x) = 1/x$ is convex ($\mathbf{dom}\, f = \{x | x > 0\}$)

Epigraph & sublevel sets

epigraph of a function f is

$$\textbf{epi}\, f = \{(x,t) \mid x \in \textbf{dom}\, f, \quad f(x) \leq t\}$$



f convex function $\Leftrightarrow$ $\textbf{epi}\, f$ convex set

the (α-)**sublevel set** of f is

$$C(\alpha) \triangleq \{x \in \textbf{dom}\, f \mid f(x) \leq \alpha\}$$

f convex $\Rightarrow$ sublevel sets are convex (converse false)

Differentiable convex functions

gradient of $f : \mathbf{R}^n \longrightarrow \mathbf{R}$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T \quad \text{(evaluated at } x\text{)}$$

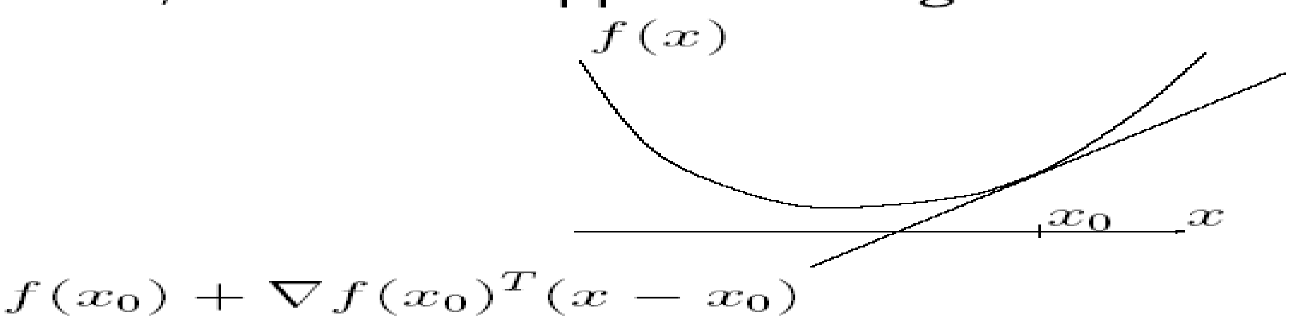
first order Taylor approximation at x_0:

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

first-order condition: for f differentiable,
f is convex $\Longleftrightarrow$ for all $x, x_0 \in \mathbf{dom}\, f$,

$$f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

i.e., 1st order approx. is a *global underestimator*



Hessian of a twice differentiable function:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

(evaluated at x)

2nd order Taylor series expansion around x_0:

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

second order condition: for f twice differentiable,
f is convex $\Longleftrightarrow$ for all $x \in \mathbf{dom}\, f$, $\nabla^2 f(x) \succeq 0$

Analytic Center

The new interior-point algorithms introduced by Karmarkar move by successive steps inside the feasible region. It is the interior of the feasible set rather than the vertices and edges that plays a dominant role in this type of algorithm. In fact, these algorithms purposely avoid the edges of the set, only eventually converging to one as a solution.

Consider a set $\mathcal{S}$ in a subset of $\mathcal{X}$ of E^n defined by a group of inequalities as

$$\mathcal{S} = \{\mathbf{x} \in \mathcal{X} : g_j(\mathbf{x}) \geqslant 0, \quad j = 1, 2, \ldots, m\},$$

and assume that the functions g_j are continuous. $\mathcal{S}$ has a nonempty interior $\mathring{\mathcal{S}} = \{\mathbf{x} \in \mathcal{X} : g_j(\mathbf{x}) > 0, \text{ all } j\}$. Associated with this definition of the set is the *potential function*

$$\psi(\mathbf{x}) = -\sum_{j=1}^{m} \log g_j(\mathbf{x})$$

defined on $\mathring{\mathcal{S}}$.

The *analytic center* of $\mathcal{S}$ is the vector (or set of vectors) that minimizes the potential; that is, the vector (or vectors) that solve

$$\min \psi(\mathbf{x}) = \min \left\{ -\sum_{j=1}^{m} \log g_j(\mathbf{x}) : \mathbf{x} \in \mathcal{X},\ g_j(\mathbf{x}) > 0 \text{ for each } j \right\}.$$

Example 1. (A cube). Consider the set $\mathcal{S}$ defined by $x_i \geqslant 0$, $(1 - x_i) \geqslant 0$, for $i = 1, 2, \ldots, n$. This is $\mathcal{S} = [0, 1]^n$, the unit cube in E^n. The analytic center can be found by differentiation to be $x_i = 1/2$, for all i. Hence, the analytic center is identical to what one would normally call the center of the unit cube.

In general, the analytic center depends on how the set is defined—on the particular inequalities used in the definition. For instance, the unit cube is also defined by the inequalities $x_i \geqslant 0$, $(1 - x_i)^d \geqslant 0$ with $d > 1$. In this case the solution is $x_i = 1/(d+1)$ for all i. For large d this point is near the inner corner of the unit cube.

Analytic Center Cont.

Let us illustrate by considering the analytic center associated with a bounded polytope Ω in E^m represented by n $(> m)$ linear inequalities; that is,

$$\Omega = \{\mathbf{y} \in E^m : \mathbf{c}^T - \mathbf{y}^T\mathbf{A} \geqslant \mathbf{0}\},$$

where $\mathbf{A} \in E^{m\times n}$ and $\mathbf{c} \in E^n$ are given and $\mathbf{A}$ has rank m. Denote the interior of Ω by

$$\mathring{\Omega} = \{\mathbf{y} \in E^m : \mathbf{c}^T - \mathbf{y}^T\mathbf{A} > \mathbf{0}\}.$$

The potential function for this set is

$$\psi_\Omega(\mathbf{y}) \equiv -\sum_{j=1}^{n} \log(c_j - \mathbf{y}^T\mathbf{a}_j) = -\sum_{j=1}^{n} \log s_j, \qquad (4)$$

where $\mathbf{s} \equiv \mathbf{c} - \mathbf{A}^T\mathbf{y}$ is a *slack vector*. Hence the potential function is the negative sum of the logarithms of the slack variables.

The analytic center of Ω is the interior point of Ω that minimizes the potential function. This point is denoted by $\mathbf{y}^a$ and has the associated $\mathbf{s}^a = \mathbf{c} - \mathbf{A}^T \mathbf{y}^a$. The pair $(\mathbf{y}^a, \mathbf{s}^a)$ is uniquely defined, since the potential function is strictly convex (see Section 7.4) in the bounded convex set Ω.

Setting to zero the derivatives of $\psi(\mathbf{y})$ with respect to each y_i gives

$$\sum_{j=1}^{n} \frac{a_{ij}}{c_j - \mathbf{y}^T \mathbf{a}_j} = 0, \text{ for all } i.$$

which can be written

$$\sum_{j=1}^{n} \frac{a_{ij}}{s_j} = 0, \text{ for all } i.$$

Now define $x_j = 1/s_j$ for each j. We introduce the notion

$$\mathbf{x} \circ \mathbf{s} \equiv (x_1 s_1, x_2 s_2, \ldots, x_n s_n)^T,$$

which is *component multiplication*. Then the analytic center is defined by the conditions

$$\begin{aligned} \mathbf{x} \circ \mathbf{s} &= \mathbf{1} \\ \mathbf{A}\mathbf{x} &= \mathbf{0} \\ \mathbf{A}^T \mathbf{y} + \mathbf{s} &= \mathbf{c}. \end{aligned}$$

The analytic center can be defined when the interior is empty or equalities are present, such as

$$\Omega = \{\mathbf{y} \in E^m : \mathbf{c}^T - \mathbf{y}^T \mathbf{A} \geqslant \mathbf{0},\ \mathbf{B}\mathbf{y} = \mathbf{b}\}.$$

In this case the analytic center is chosen on the linear surface $\{\mathbf{y} : \mathbf{B}\mathbf{y} = \mathbf{b}\}$ to maximize the product of the slack variables $\mathbf{s} = \mathbf{c} - \mathbf{A}^T\mathbf{y}$. Thus, in this context the interior of Ω refers to the interior of the positive orthant of slack variables: $R_+^n \equiv \{\mathbf{s} :\ \mathbf{s} \geqslant \mathbf{0}\}$. This definition of interior depends only on the region of the slack variables. Even if there is only a single point in Ω with $\mathbf{s} = \mathbf{c} - \mathbf{A}^T\mathbf{y}$ for some $\mathbf{y}$ where $\mathbf{B}\mathbf{y} = \mathbf{b}$ with $\mathbf{s} > \mathbf{0}$, we still say that $\mathring{\Omega}$ is not empty.

Central Path

Consider a primal linear program in standard form

$$\text{(LP) minimize } \mathbf{c}^T\mathbf{x} \quad (5)$$
$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{x} \geqslant \mathbf{0}.$$

We denote the feasible region of this program by $\mathcal{F}_p$. We assume that $\mathring{\mathcal{F}}_p = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} > \mathbf{0}\}$ is nonempty and the optimal solution set of the problem is bounded.

Associated with this problem, we define for $\mu \geqslant 0$ the *barrier problem*

$$\text{(BP) minimize } \mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^{n} \log x_j \quad (6)$$
$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{x} > \mathbf{0}.$$

It is clear that $\mu = 0$ corresponds to the original problem (5). As $\mu \to \infty$, the solution approaches the analytic center of the feasible region (when it is bounded), since the barrier term swamps out $\mathbf{c}^T\mathbf{x}$ in the objective. As μ is varied continuously toward 0, there is a path $\mathbf{x}(\mu)$ defined by the solution to (BP). This path $\mathbf{x}(\mu)$ is termed the *primal central path*. As $\mu \to 0$ this path converges to the analytic center of the optimal face $\{\mathbf{x} : \mathbf{c}^T\mathbf{x} = z^*, \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$, where z^* is the optimal value of (LP).

A strategy for solving (LP) is to solve (BP) for smaller and smaller values of μ and thereby approach a solution to (LP). This is indeed the basic idea of interior-point methods.

At any $\mu > 0$, under the assumptions that we have made for problem (5), the necessary and sufficient conditions for a unique and bounded solution are obtained by introducing a *Lagrange multiplier* vector $\mathbf{y}$ for the linear equality constraints to form the *Lagrangian* (see Chapter 11)

$$\mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^{n} \log x_j - \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}).$$

$$\mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^{n} \log x_j - \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}).$$

The derivatives with respect to the x_j's are set to zero, leading to the conditions

$$c_j - \mu/x_j - \mathbf{y}^T\mathbf{a}_j = 0, \text{ for each } j$$

or equivalently

$$\mu\mathbf{X}^{-1}\mathbf{1} + \mathbf{A}^T\mathbf{y} = \mathbf{c} \tag{7}$$

where as before $\mathbf{a}_j$ is the j-th column of $\mathbf{A}, \mathbf{1}$ is the vector of 1's, and $\mathbf{X}$ is the diagonal matrix whose diagonal entries are the components of $\mathbf{x} > \mathbf{0}$. Setting $s_j = \mu/x_j$ the complete set of conditions can be rewritten

$$\begin{aligned} \mathbf{x} \circ \mathbf{s} &= \mu\mathbf{1} \\ \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{A}^T\mathbf{y} + \mathbf{s} &= \mathbf{c}. \end{aligned} \tag{8}$$

Central path

Example 2. (A square primal). Consider the problem of maximizing x_1 within the unit square $\mathcal{S} = [0, 1]^2$. The problem is formulated as

$$\begin{array}{ll} \min & -x_1 \\ \text{subject to} & x_1 + x_3 = 1 \\ & x_2 + x_4 = 1 \\ & x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_3 \geqslant 0,\ x_4 \geqslant 0. \end{array}$$

Here x_3 and x_4 are slack variables for the original problem to put it in standard form. The optimality conditions for $\mathbf{x}(\mu)$ consist of the original 2 linear constraint equations and the four equations

$$\begin{aligned} y_1 + s_1 &= 1 \\ y_2 + s_2 &= 0 \\ y_1 + s_3 &= 0 \\ y_2 + s_4 &= 0 \end{aligned}$$

together with the relations $s_i = \mu/x_i$ for $i = 1, 2 \ldots, 4$. These equations are readily solved with a series of elementary variable eliminations to find

$$x_1(\mu) = \frac{1 - 2\mu \pm \sqrt{1 + 4\mu^2}}{2}$$

$$x_2(\mu) = 1/2.$$

Using the "+" solution, it is seen that as $\mu \to 0$ the solution goes to $\mathbf{x} \to (1, 1/2)$. Note that this solution is not a corner of the cube. Instead it is at the analytic center of the optimal face $\{\mathbf{x} : x_1 = 1,\ 0 \leqslant x_2 \leqslant 1\}$. See Fig. 5.2. The limit of $\mathbf{x}(\mu)$ as $\mu \to \infty$ can be seen to be the point $(1/2, 1/2)$. Hence, the central path in this case is a straight line progressing from the analytic center of the square (at $\mu \to \infty$) to the analytic center of the optimal face (at $\mu \to 0$).

Dual central path

Now consider the dual problem

$$\begin{array}{lll}
\text{(LD)} & \text{maximize} & \mathbf{y}^T\mathbf{b} \\
& \text{subject to} & \mathbf{y}^T\mathbf{A} + \mathbf{s}^T = \mathbf{c}^T \\
& & \mathbf{s} \geq \mathbf{0}.
\end{array}$$

We may apply the barrier approach to this problem by formulating the problem

$$\begin{array}{lll}
\text{(BD)} & \text{maximize} & \mathbf{y}^T\mathbf{b} + \mu \sum_{j=1}^{n} \log s_j \\
& \text{subject to} & \mathbf{y}^T\mathbf{A} + \mathbf{s}^T = \mathbf{c}^T \\
& & \mathbf{s} > \mathbf{0}.
\end{array}$$

We assume that the dual feasible set $\mathcal{F}_d$ has an interior $\mathring{\mathcal{F}}_d = \{(\mathbf{y}, \mathbf{s}) : \mathbf{y}^T\mathbf{A} + \mathbf{s}^T = \mathbf{c}^T,\ \mathbf{s} > \mathbf{0}\}$ is nonempty and the optimal solution set of (LD) is bounded. Then, as μ is varied continuously toward 0, there is a path $(\mathbf{y}(\mu), \mathbf{s}(\mu))$ defined by the solution to (BD). This path $\mathbf{y}(\mu)$ is termed the *dual central path*.

To work out the necessary and sufficient conditions we introduce $\mathbf{x}$ as a Lagrange multiplier and form the Lagrangian

$$\mathbf{y}^T\mathbf{b} + \mu \sum_{j=1}^{n} \log s_j - (\mathbf{y}^T\mathbf{A} + \mathbf{s}^T - \mathbf{c}^T)\mathbf{x}.$$

Setting to zero the derivative with respect to y_i leads to

$$b_i - \mathbf{a}_{i.}\mathbf{x} = 0, \text{ for all } i$$

where $\mathbf{a}_{.i}$ is the i-th row of A. Setting to zero the derivative with respect to s_j leads to

$$\mu/s_j - x_j = 0, \text{ for all } j.$$

Combining these equations and including the original constraint yields the complete set of conditions

$$\begin{aligned} \mathbf{x} \circ \mathbf{s} &= \mu \mathbf{1} \\ \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{A}^T\mathbf{y} + \mathbf{s} &= \mathbf{c}. \end{aligned}$$

Example 3. (The square dual). Consider the dual of example 2. This is

$$\begin{aligned} \max \quad & y_1 + y_2 \\ \text{subject to} \quad & y_1 \leqslant -1 \\ & y_2 \leqslant 0. \end{aligned}$$

(The values of s_1 and s_2 are the slack variables of the inequalities.) The solution to the dual barrier problem is easily found from the solution of the primal barrier problem to be

$$y_1(\mu) = -1 - \mu/x_1(\mu), \quad y_2 = -2\mu.$$

As $\mu \to 0$, we have $y_1 \to -1$, $y_2 \to 0$, which is the unique solution to the dual LP. However, as $\mu \to \infty$, the vector y is unbounded, for in this case the dual feasible set is itself unbounded.

Primal–Dual Central Path

Suppose the feasible region of the primal (LP) has interior points and its optimal solution set is bounded. Then, the dual also has interior points (see Exercise 4). The primal–dual path is defined to be the set of vectors $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ that satisfy the conditions

$$\begin{aligned} \mathbf{x} \circ \mathbf{s} &= \mu \mathbf{1} \\ \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{A}^T\mathbf{y} + \mathbf{s} &= \mathbf{c} \\ \mathbf{x} \geqslant \mathbf{0}, \;\; \mathbf{s} &\geqslant \mathbf{0} \end{aligned} \tag{9}$$

Proposition 1. *Suppose the feasible sets of the primal and dual programs contain interior points. Then the primal–dual central path* $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ *exists for all* μ, $0 \leqslant \mu < \infty$. *Furthermore,* $\mathbf{x}(\mu)$ *is the primal central path, and* $(\mathbf{y}(\mu), \mathbf{s}(\mu))$ *is the dual central path. Moreover,* $\mathbf{x}(\mu)$ *and* $(\mathbf{y}(\mu), \mathbf{s}(\mu))$ *converge to the analytic centers of the optimal primal solution and dual solution faces, respectively, as* $\mu \to 0$.