Notes on the Khachiyan-Kalantari algorithm

Vašek Chvátal

1 Linear programming reformulated

The problem of testing solvability of systems of linear inequalities is polynomially reducible to the problem of finding nonnegative nonzero solutions of systems of homogeneous linear equations. To see this, suppose that we have an oracle for solving the latter problem; we are going to describe a way of using this oracle to solve the former problem.

A system

$$Ax \le b \tag{1}$$

of m linear inequalities is unsolvable if and only if it is inconsistent in the sense that the system

$$y \ge 0,\ y^T A = 0,\ y^T b < 0 \tag{2}$$

in m variables has a solution. If the system

$$y^T A = 0,\ y^T b + s = 0 \tag{3}$$

of homogeneous linear equations in $m+1$ variables has no nonnegative nonzero solution, then (2) has no solution, and so (1) is solvable; if the oracle finds a nonnegative nonzero solution y, s of (3), then we distinguish between two cases. In case $s > 0$, system (2) has a solution, and so (1) is unsolvable; in case $s = 0$, we have a nonzero vector y such that

$$y \ge 0,\ y^T A = 0,\ y^T b = 0$$

and we will use this vector to reduce the size of (1).

Writing (1) in the extensive form as

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = 1, 2, \ldots, m),$$

note that every solution of (1) must satisfy

$$0 = \sum_{j=1}^{n} \left(\sum_{i=1} y_i a_{ij} \right) x_j = \sum_{i=1} y_i \left(\sum_{j=1}^{n} a_{ij} x_j \right) \leq \sum_{i=1} y_i b_i = 0,$$

and so it must satisfy

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad \text{whenever } y_i > 0.$$

Finally, consider an arbitrary subscript i such that $y_i > 0$. If $a_{ik} \neq 0$ for some k, then we may eliminate x_k from (1) by the substitution

$$x_k = b_i / a_{ik} - \sum_{j \neq k} (a_{ij}/a_{ik})\, x_j \;;$$

if $a_{ij} = 0$ for all j, then we may either (in case $b_i \geq 0$) reduce (1) simply by deleting the i-th inequality or (in case $b_i < 0$) conclude at once that (1) is unsolvable.

2 Diagonal matrix scaling

A matrix Q in $\mathbf{R}^{n \times n}$ is called *positive semidefinite* if $x^T Q x \geq 0$ for all vectors x in $\mathbf{R}^n$. We let e denote the vector $[1, 1, \ldots, 1]^T$ in $\mathbf{R}^n$.

THEOREM 1 *Every symmetric positive semidefinite matrix Q has precisely one of the following two properties:*

(i) *there is a diagonal matrix D such that $De > 0$ and $(DQD)e = e$,*

(ii) *there is a nonnegative nonzero vector x such that $Qx = 0$.* □

To see that no symmetric matrix Q has both of these properties, assume the contrary; now $0 = (Qx)^T (De) = x^T (QDe) = x^T (D^{-1} e) > 0$, a contradiction. In Section 5, we shall prove that every symmetric positive semidefinite matrix Q has at least one of these properties.

Theorem 1 relates to linear programming as follows. Given an arbitrary matrix A in $\mathbf{R}^{m\times n}$, write $Q = A^T A$ and observe that Q is symmetric positive semidefinite. If x is a vector such that $Qx = 0$, then $0 = x^T Q x = (Ax)^T(Ax)$, and so $Ax = 0$; in particular, if Q has property (ii), then the nonnegative nonzero vector x satisfies $Ax = 0$. If Q has property (i), then write $y = ADe$ and observe that $A^T y = QDe = D^{-1}e > 0$; now the system $Ax = 0$ can have no nonnegative nonzero solution x since any such x would satisfy $0 = (Ax)^T y = x^T(A^T y) > 0$, a contradiction.

3 The algorithm

The Khachiyan-Kalantari algorithm, given a symmetric positive semidefinite $n\times n$ matrix A and positive numbers δ, ε less than 1, returns either a diagonal matrix D such that

$$De > 0 \quad \text{and} \quad \|DADe - e\| < \delta$$

or a vector x such that

$$\|x\| = 1 \quad \text{and} \quad x^T Ax < \varepsilon.$$

In the description of the algorithm, I denotes the identity matrix and $\operatorname{diag}(v)$ denotes the diagonal matrix whose diagonal is v.

$\rho = (1 - 1/(1 + 4\sqrt{n}))^{1/2}$;
$D_0 = I$, $k = 0$;
while $(D_k e)^T A(D_k e) \geq \varepsilon \cdot \|D_k e\|^2$ and $\|D_k A D_k e - e\| \geq 3/4$
do solve the system $(I + D_k A D_k)z = e - D_k A D_k e - \rho^k D_k(e - Ae)$;
$D_{k+1} = \rho \cdot \operatorname{diag}(D_k(e+z))$, $k = k+1$;
end
if $(D_k e)^T A(D_k e) < \varepsilon \cdot \|D_k e\|^2$
then **return** the vector $\|D_k e\|^{-1} \cdot D_k e$;
else $D_0 = D_k$, $k = 0$;
while $\|D_k A D_k e - e\| \geq \delta$
do solve the system $(I + D_k A D_k)z = e - D_k A D_k e$;
$D_{k+1} = \operatorname{diag}(D_k(e+z))$, $k = k+1$;
end
return the matrix D_k;
end

It is not immediately obvious that the algorithm terminates. In Section 4, we shall prove that it does; in fact, we shall give the following upper bound on the number of its iterations.

THEOREM 2 *In the Khachiyan-Kalantari algorithm, the first* **while** *loop goes through at most*

$$\lceil (1+4\sqrt{n}) \cdot \ln((16n+4\sqrt{n}+4)\cdot \|e-Ae\|^2 \cdot \varepsilon^{-1})\rceil$$

iterations and the second **while** *loop goes through at most*

$$\lceil \lg\lg(1/\delta) - \lg\lg(4/3)\rceil$$

iterations.

The second **while** loop of the Khachiyan-Kalantari algorithm is an application of *Newton's method*. This general method, given a mapping $F : \mathbf{R}^n \to \mathbf{R}^n$, constructs a sequence of points in $\mathbf{R}^n$ aimed at approximating a solution x of $F(x) = 0$. For each point x in this sequence, the method constructs the next point $x + y$ by finding a solution y of

$$F(x) + J(x)y = 0,$$

where $J(x)$ is the matrix featured in the linear approximation

$$F(x+y) \;\approx\; F(x) + J(x)y.$$

This matrix is called the *Jacobian matrix*; the entry in its i-th row and its j-th column equals

$$\frac{\partial F_i}{\partial x_j}(x),$$

where $F_i(x)$ denotes the i-th component of $F(x)$ and x_j denotes the j-th component of x. In the Khachiyan-Kalantari application, F is defined by

$$F(x) = ADe - D^{-1}e \;\text{ with }\; D = \mathrm{diag}(x):$$

for this choice of F, we have

$$J(x) = A + D^{-2},$$

and so the system $F(x) + J(x)y = 0$ can be solved by setting

$$y = Dz \;\text{ with }\; (I + DAD)z = e - DADe.$$

As we shall prove later, the initial condition $\|D_0AD_0e - e\| < 3/4$ guarantees a doubly exponential decrease of $\|D_kAD_ke - e\|$.

The first **while** loop falls in the category of *path-following methods*, also called *homotopy methods*. These methods, given a mapping $G : \mathbf{R}^n \to \mathbf{R}^n$, construct a sequence of points in $\mathbf{R}^n$ aimed at approximating a solution x of $G(x) = 0$. For this purpose, they define a mapping $H : \mathbf{R}^n \times [0, 1] \to \mathbf{R}^n$ such that

(i) a solution x of $H(x, 1) = 0$ is readily available and

(ii) $H(x, 0) = G(x)$ for all x;

then they construct a sequence of points (x, t) in $\mathbf{R}^n \times [0, 1]$, with t starting at 1 and monotonically converging to 0, that approximate solutions of

$$\{(x, t) : H(x, t) = 0\}.$$

In the Khachiyan-Kalantari algorithm, H is defined by

$$H(x, t) = DADe - e + tD(e - Ae) \;\text{ with }\; D = \mathrm{diag}(x);$$

as we shall prove later, the first **while** loop maintains the invariant

$$\|H(D_ke,\, \rho^k)\| < \frac{1}{2}\,.$$

Each iteration of this **while** loop represents an iteration of Newton's method: if

$$F(x) = ADe - D^{-1}e + \rho^k(e - Ae) \;\text{ with }\; D = \mathrm{diag}(x),$$

then $J(x) = A + D^{-2}$, the system $F(D_ke) + J(D_ke)y = 0$ can be solved by setting

$$y = D_kz \;\text{ with }\; (I + D_kAD_k)z = e - D_kAD_ke - \rho^kD_k(e - Ae),$$

and so the transition $D_k \mapsto \rho^{-1}D_{k+1}$ is a Newton step. (A change of variable, $N_k = \rho^{-k}D_k$, makes this transition simply $N_k \mapsto N_{k+1}$.)

4 Proof of Theorem 2

LEMMA 1 *If Q is a positive semidefinite matrix in $\mathbf{R}^{n\times n}$ and if z is a vector in $\mathbf{R}^n$, then $\|(I+Q)z\| \geq \|z\|$.*

Proof. Schwarz's inequality (also called the Cauchy-Schwarz inequality or — after its original discoverer — the Buniakovskii inequality) asserts that $|x^T y| \leq \|x\| \cdot \|y\|$. In particular,

$$\|z\| \cdot \|(I+Q)z\| \geq z^T(I+Q)z = \|z\|^2 + z^T Q z \geq \|z\|^2.$$

□

LEMMA 2 *Let A be a positive semidefinite matrix, let D be a diagonal matrix, let z and b be vectors such that*

$$(I+DAD)z = e - DADe - Db$$

and let D' be the matrix defined by

$$D' = \mathrm{diag}(D(e+z)).$$

Then

(i) $\|e - D'AD'e - D'b\| \leq \|e - DADe - Db\|^2$,

(ii) *if $\|e - DADe - Db\| < 1$ and $De > 0$, then $D'e > 0$.*

Proof. Writing $Z = \mathrm{diag}(z)$, observe that $D' = (I+Z)D$, and so

$$e - D'AD'e - D'b = e - (I+Z)DAD(e+z) - (I+Z)Db = Zz.$$

We have

$$\|Zz\| = (\textstyle\sum_j z_j^4)^{1/2} \leq (\textstyle\sum_j z_j^2) = \|z\|^2$$

and, by Lemma 1 with $Q = DAD$,

$$\|z\| \leq \|e - DADe - Db\|. \tag{4}$$

This proves (i). If $\|e - DADe - Db\| < 1$, then (4) guarantees that $\|z\| < 1$, and so $e + z > 0$; this inequality and $De > 0$ imply $D'e > 0$. □

LEMMA 3 *The first* **while** *loop maintains the invariant*

$$\|e - D_k A D_k e - \rho^k D_k(e - Ae)\| < \frac{1}{2}\ . \tag{5}$$

Proof. By induction on k. If $k = 0$, then (5) is satisfied as its left-hand side equals zero. If (5) is satisfied for some value of k, then (i) of Lemma 2 with $D = D_k$, $b = \rho^k(e - Ae)$, and $D' = \rho^{-1} D_{k+1}$ guarantees that

$$\|e - \rho^{-2} D_{k+1} A D_{k+1} e - \rho^{k-1} D_{k+1}(e - Ae)\| < \frac{1}{4}\ . \tag{6}$$

In turn, (6) implies that

$$\|\rho^{-2} D_{k+1} A D_{k+1} e + \rho^{k-1} D_{k+1}(e - Ae)\| < \frac{1}{4} + \|e\| = \frac{1 + 4\sqrt{n}}{4}\ ;$$

multiplying both sides of this inequality by $1 - \rho^2$, we get

$$\|(1 - \rho^{-2}) D_{k+1} A D_{k+1} e + (\rho^{k+1} - \rho^{k-1}) D_{k+1}(e - Ae)\| < \frac{1}{4}\ .$$

The sum of this inequality and (6) shows that (5) holds with $k + 1$ in place of k. □

LEMMA 4 *The first* **while** *loop maintains the invariant*

$$D_k e > 0\ . \tag{7}$$

Proof. Invariant (5) and (ii) of Lemma 2 with $D = D_k$, $b = \rho^k(e - Ae)$, and $D' = \rho^{-1} D_{k+1}$. □

LEMMA 5 *The first* **while** *loop maintains the invariant*

$$\frac{(D_k e)^T A (D_k e)}{\|D_k e\|^2} < \left(16n + 4\sqrt{n} + 4\right) \|e - Ae\|^2 \cdot \rho^{2k}\ . \tag{8}$$

Proof. Invariant (5) and the condition $\|D_k A D_k e - e\| \geq 3/4$ guarantee that

$$\|\rho^k D_k(e - Ae)\| > \frac{1}{4}\ ; \tag{9}$$

since D_k is a diagonal matrix, we have

$$\|D_k(e - Ae)\| \le \|D_k e\| \cdot \|e - Ae\| \; ;$$

this inequality and (9) imply

$$\|D_k e\| > \frac{1}{4\|e - Ae\|\rho^k} \; . \tag{10}$$

Schwarz's inequality and invariant (5) guarantee that

$$e^T(D_k A D_k e + \rho^k D_k(e - Ae) - e) \le \frac{\sqrt{n}}{4} \; ,$$

and so

$$(D_k e)^T A (D_k e) \le \frac{\sqrt{n}}{4} - \rho^k (D_k e)^T (e - Ae) + e^T e \; ,$$

which, by Schwarz's inequality once again, implies

$$(D_k e)^T A (D_k e) \le \frac{\sqrt{n}}{4} + \rho^k \|D_k e\| \cdot \|(e - Ae)\| + n \; .$$

This inequality and (10) imply (8). □

LEMMA 6 *The second* **while** *loop maintains the invariant*

$$D_k > 0 \;\; \textit{and} \;\; \|D_k A D_k e - e\| \le \left(\frac{3}{4}\right)^{2^k} .$$

Proof. By induction on k, using Lemma 2 with $D = D_k$, $b = 0$, and $D' = D_{k+1}$. □

5 Proof of Theorem 1

We are going to prove that every symmetric positive semidefinite matrix Q has at least one of the properties

(i) there is a diagonal matrix D such that $De > 0$ and $(DQD)e = e$ and

(ii) there is a nonnegative nonzero vector x such that $Qx = 0$.

For this purpose, let n denote the order of Q; write

$$S_+ = \{x \in \mathbf{R}^n : x \ge 0, \|x\| = 1\}$$

and $f(x) = x^T Qx$ for all x in $\mathbf{R}^n$. Since f is a continuous function and since S_+ is a compact set, there is a point x^* that minimizes f over S_+; since Q is positive semidefinite, $f(x^*) \geq 0$; we shall distinguish between two cases.

Case 1: $f(x^*) > 0$.

Let us write $\mathbf{R}^n_+ = \{x \in \mathbf{R}^n : x > 0\}$ and let us prove that the function

$$g : \mathbf{R}^n_+ \to \mathbf{R}$$

defined by

$$g(x) = \frac{1}{2} x^T Q x - \sum_{j=1}^{n} \ln x_j$$

has the following properties:

(a) for every positive t there is a positive r such that $g(x) > t$ whenever $x \in \mathbf{R}^n_+$ and $\|x\| > r$,

(b) for every positive t there is a positive ε such that $g(x) > t$ whenever $x \in \mathbf{R}^n_+$ and $\min_j x_j < \varepsilon$.

For this purpose, note first that

$$g(x) \geq \frac{1}{2} f(x^*) \|x\|^2 - \ln(\min_j x_j) - (n-1) \cdot \ln \|x\| \geq \frac{1}{2} f(x^*) \|x\|^2 - n \cdot \ln \|x\|$$

for all x in $\mathbf{R}^n_+$; now (a) follows from the asumption of this case and (b) follows in turn.

Since g is continuous and differentiable, (a) and (b) guarantee that g attains its minimum over $\mathbf{R}^n_+$ and that every x minimizing g over $\mathbf{R}^n_+$ satisfies

$$\frac{\partial g}{\partial x_j}(x) = 0 \text{ for all } j = 1, 2, \ldots, n$$

(actually, a little more careful inspection shows that g is strictly convex, and so x is unique); in terms of $D = \operatorname{diag}(x)$, this system reads $QDe - D^{-1}e = 0$, and so Q has property (i).

CASE 2: $f(x^*) = 0$.

By assumption of this case, x^* minimizes f over $\mathbf{R}^n$, and so

$$\frac{\partial f}{\partial x_j}(x) = 0 \ \text{ for all } j = 1, 2, \ldots, n,$$

which means $Qx^* = 0$, and so Q has property (ii). $\square$

Reference

L. Khachiyan and B. Kalantari, Diagonal matrix scaling and linear programming. *SIAM J. Optim.* **2** (1992), 668–672.