The Maxdet Problem – Algorithm and Applications in Machine Learning

Yu Xia
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu
Minato-Ku, Tokyo

Tokyo 106-8569, Japan
yuxia@ism.ac.jp

Abstract

We show that the problem of maximizing a sum of linear functions and weighted logarithmic determinants with linear positive semidefinite matrix constraints has some interesting applications in machine learning. We extend the primal-dual interior-point path-following algorithm to the problem. Polynomial-time complexity of the algorithm is also presented.

Keywords: semidefinite programming, interior point method, machine learning, complexity.

1 Introduction

Denote the vector space of real $m \times m$ symmetric matrices as $\mathcal{S}^m$. Since the cone of positive semidefinite matrices induces a partial order, for $X \in \mathcal{S}^m$, we use $X \succeq 0$ ($\succ 0$) to represent that X is positive semidefinite (positive definite). Let $A \bullet B$ denote the standard inner product on $\mathcal{S}^m$:

$$A \bullet B = \operatorname{tr} AB = \sum_{1 \le i,j \le m} A_{ij} B_{ij} \ .$$

We use $\operatorname{Det} X$ to represent the determinant of a matrix X. The nature logarithmic function is denoted by $\ln(\cdot)$.

Our variables are n symmetric matrices with dimensions $N_1, \dots, N_n > 0$. Let LD denote a subset of the index set $\{1, \dots, n\}$.

We consider the following determinant maximization problem with linear positive semidefinite matrix constraints (maxdet-problem).

$$\begin{array}{ll} \min_{X_i \in \mathcal{S}^{N_i}} & \sum_{i=1}^n C_i \bullet X_i - \sum_{j \in \mathrm{LD}} \hat{c}_j \ln \operatorname{Det} X_j \\ \text{s.t.} & \sum_{i=1}^n A_{ki} \bullet X_i = b_k \quad (k = 1, \dots, m) \ , \\ & X_i \succeq 0 \, (i \notin \mathrm{LD}) \ , \quad X_i \succ 0 \, (i \in \mathrm{LD}) \ . \end{array} \tag{1}$$

Here, $C_i, A_{ki} \in \mathcal{S}^{N_i}$, $\hat{c}_j > 0$, $\mathbf{b} = (b_1, \dots, b_m)^T$ are given data. Problem (1) is a convex program. Since the derivative of $(\ln \operatorname{Det} X)$ is X^{-1}, it is easy to see that the dual to (1) is

$$\begin{array}{ll} \max_{\mathbf{y} \in \mathbb{R}^m, Z_i \in \mathcal{S}^{N_i}} & \mathbf{b}^T \mathbf{y} + \sum_{j \in \mathrm{LD}} \hat{c}_j \ln \operatorname{Det} Z_j + \sum_{j \in \mathrm{LD}} \hat{c}_j N_j \ln \hat{c}_j \\ \text{s.t.} & \sum_{k=1}^m y_k A_{ki} + Z_i = C_i \quad (i = 1, \dots, n) \ , \\ & Z_i \succeq 0 \ (i \notin \mathrm{LD}) \ , \quad Z_i \succ 0 \ (i \in \mathrm{LD}) \ . \end{array} \tag{2}$$

Maxdet-problem is an extension of the semidefinite programming (SDP) problem [1]. It includes the analytic centering problem as well. Vandenberghe et. al in [11] give many applications of the maxdet-problem in computational geometry, statistics, information and communication theory, etc. In [11], an interior-point algorithm for the maxdet-problem based on self-concordant barrier [7] is given with numerical examples. Toh presents a more general algorithmic framework based on a symmetrized Newton equation for this problem in [9], which includes the search direction of [11]. Toh shows that the algorithms are efficient, robust and accurate through numerical results.

In this paper, we first show some applications of the maxdet-problem in machine learning, then introduce a primal-dual path-following interior-point algorithm for the maxdet-problem based on symmetrized Newton equations.

The rest of the paper is organized as follows. In § 2, we give some applications the maxdet-problem in machine learning. In § 3, we introduce our primal-dual interior point path-following algorithm for the maxdet-problem. Finally, in § 4, we summarize complexity results of our algorithm.

2 Applications in Machine Learning

In this part, we give some applications of the maxdet-problem in machine learning.

There is a system with input $\mathbf{x}$ and output $\mathbf{y}$. A *training set* of n data items $(\mathbf{x}_i, \mathbf{y}_i)$, $(i = 1, \ldots, n)$ is given. The task is to predict output $\mathbf{y}$ for new input $\mathbf{x}$. The prediction rule is often described by some model. And the parameters of the model can be estimated by the training data.

Maximum likelihood estimation is a principle for estimation. It assumes that the most reasonable values of the parameters are those for which the probability of the cases represented by the training data is the largest.

2.1 Density Estimation

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ be a random sample of independent identically-distributed (i.i.d.) observation from a probability density $f(\mathbf{x})$. Suppose f is *multivariate Gaussian* with *mean* $\boldsymbol{\mu}$ *covariance* Σ:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}(\det \Sigma)^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] .$$

The task is to estimate $\boldsymbol{\mu}$ and Σ based on the data. The log-probability of the sample is

$$L(\boldsymbol{\mu}, \Sigma) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\det\Sigma - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})$$

Define the empirical mean and empirical covariance:

$$\bar{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \qquad \bar{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\boldsymbol{\mu}})(\mathbf{x}_i-\bar{\boldsymbol{\mu}})^T .$$

Then the log-probability can be represented as

$$L(\boldsymbol{\mu}, \Sigma) = -\frac{n}{2}\left[p\ln(2\pi) + \ln\det\Sigma + \operatorname{tr}(\Sigma^{-1}\bar{\Sigma}) + (\boldsymbol{\mu}-\bar{\boldsymbol{\mu}})^T\Sigma^{-1}(\boldsymbol{\mu}-\bar{\boldsymbol{\mu}})\right] .$$

The mean can be estimated by $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}$ to maximize $L(\boldsymbol{\mu}, \Sigma)$.

The likelihood equations for the positive definte covariance matrix Σ is

$$\Sigma^{-1} = \Sigma^{-1}\bar{\Sigma}\Sigma^{-1} \ .$$

When $n \geq p+1$, $\bar{\Sigma}$ is positive definite with probability one.

When $n < p+1$, or there is some prior information of the structure of Σ, to estimate Σ, some linear constraints may be imposed on Σ. For instance, set some entries zero, or add bounds on some variances $\Sigma_{ii} \leq a$, which can be represented as

$$\begin{pmatrix} \Sigma^{-1} & e_i \\ e_i^T & a \end{pmatrix} \succeq \mathbf{0} \ .$$

Let $M \stackrel{\text{def}}{=} \Sigma^{-1}$. Then the maximum likelihood estimation of the density function f can be represented as a maxdet-problem. See [11].

$$\begin{array}{ll} \max & -\operatorname{tr}(M\bar{\Sigma}) + \ln\det M \\ \text{s.t.} & M \in V \\ & M \succ \mathbf{0} \ , \end{array}$$

where V is the feasible set represented by the linear constraints.

The density estimation problem can also be modeled as a maxdet-problem if the data is one-dimentional and the density model is assumed to be

$$f(x;\alpha,\beta) = p(x;\alpha)K(x;\beta) \ ,$$

where $p(x;\alpha)$ is a polynomial in x with parameter α and $K(x;\beta)$ is a density function of x with parameter β; see [3].

2.2 Classification

If the output $\mathbf{y}$ of the system are descriptive labels (also referred to as *classes*, categories), the prediction task is called *classification*.

Bayes classifier is a solution to classification. The rule is to categorize the data to the most probable class based on the conditional (discrete) distribution $\Pr(\mathbf{y}|\mathbf{x})$.

Denote the possible classes as $\{1,\ldots,q\}$. Let $g_k(\mathbf{x})$ be the calss-conditional density of $\mathbf{x}$ in class $\mathbf{y} = k$. Let π_k denote the prior probability of class k, such that $\sum_{k=1}^q \pi_k = 1$.

By *Bayes theorem*,

$$\Pr(\mathbf{y} = k | \mathbf{x} = \tilde{\mathbf{x}}) = \frac{g_k(\tilde{\mathbf{x}})\pi_k}{\sum_{i=1}^q g_i(\tilde{\mathbf{x}})\pi_i} \ .$$

Suppose the density of each class is multivariate Gaussian. The *quadratic discriminant function (QDA)* for class k is

$$f_k(\mathbf{x}) = -\frac{1}{2}\ln\det\Sigma_k - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^T\Sigma_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k) + \ln\pi_k \ .$$

The decision boundary between classes k and l is

$$\{\mathbf{x}\colon g_k(\mathbf{x}) = g_l(\mathbf{x})\} \ .$$

See [5].

Let n_k be the number of observations $\mathbf{y} = k$ in the training set. Using maximum likelihood estimation,

$$\hat{\pi}_k = \frac{n_k}{n} \ .$$

And, QDA can be obtained by the technique described in the *density estimation* part.

2.3 Entropy Approximation

The *differential entropy* [8] of a random variable $\mathbf{x}$ with density $g(\mathbf{x})$ is defined as

$$H(\mathbf{x}) = -\int g(\mathbf{x}) \ln g(\mathbf{x}) d\mathbf{x} \ ,$$

which is a measure of uncertainty of the probability density. Entropies are usually difficult to calculate. The differential entropy of a multivariate Gaussian distribution is

$$H_l(\mathbf{x}) = \frac{1}{2} \left[p + p \ln(2\pi) + \ln \det \Sigma \right] \ .$$

As is discussed before, H_l with some additional constraints can be formulated as a maxdet-problem. It is known that Gaussian density has the maximum entropy among all density functions with the same covariance. So we can obtain an upper bound of the entropy of a density distribution. In some algorithm for entropy calculation, the density is assumed to be near Gaussian and approximated by H_l.

3 The Path-Following Algorithm

Because the coefficients for the determinants are fixed, some properties of SDP that are essential to its convergence analysis are no longer valid for the maxdet-problem. For instance, the primal and dual search directions are not necessarily orthogonal in the neighborhood of the central path defined in [9]. To overcome this difficulty, we introduce *the biased center*. We also develop *the adaptive set strategy* to deal with the cases when the coefficients of the determinants are smaller than the central path parameter. See [10] for details.

We describe the central path of the maxdet problem in §§3.1, derive the Newton system on the central path in §§3.2, define the neighborhood of the central path, and finally present the interior-point path-following algorithm for the maxdet problem in §§3.3.

3.1 The Central Path

To describe the interior-point path-following algorithm for the maxdet-problem, we first deduce the central path for (1)-(2).

It is not hard to see that there is no duality gap between (1) and (2) iff the following conditions are satisfied: i) $(X, \mathbf{y}, Z)$ are feasible; and ii)

$$X_i \bullet Z_i = 0 \ (i \notin \mathrm{LD}) \ , \quad X_i Z_i = \hat{c}_i I \ (i \in \mathrm{LD}) \ .$$

Applying the logarithmic barrier functions to the primal and dual problems (1)-(2), we come to the following parameterized primal-dual pair:

$$\begin{array}{lll} \min_{X_i \in \mathcal{S}^{N_i}} & \sum_{i=1}^n C_i \bullet X_i - \sum_{j \in \mathrm{LD}} \hat{c}_j \ln \mathrm{Det}\, X_j - \mu \sum_{j \notin \mathrm{LD}} \ln \mathrm{Det}\, X_j & \\ \text{s.t.} & \sum_{i=1}^n A_{ki} \bullet X_i = b_k \quad (k = 1, \ldots, m) \ , & (3) \\ & X_i \succ 0 \quad (i = 1, \ldots, n) \ . & \end{array}$$

$$\begin{array}{lll} \max_{\substack{\mathbf{y} \in \mathbb{R}^m \\ Z_i \in \mathcal{S}^{N_i}}} & \mathbf{b}^T \mathbf{y} + \sum_{j \in \mathrm{LD}} \hat{c}_j \ln \mathrm{Det}\, Z_j + \sum_{j \in \mathrm{LD}} \hat{c}_j N_j + \mu \sum_{j \notin \mathrm{LD}} \ln \mathrm{Det}\, Z_j & \\ \text{s.t.} & \sum_{k=1}^m y_k A_{ki} + Z_i = C_i \quad (i = 1, \ldots, n) \ , & (4) \\ & Z_i \succ 0 \quad (i = 1, \ldots, n) \ . & \end{array}$$

Let $N \stackrel{\text{def}}{=} \sum_{i=1}^n N_i$. Define the operator $\mathcal{A}\colon \mathcal{S}^{N_1} \times \cdots \times \mathcal{S}^{N_n} \to \mathbb{R}^m$ as

$$(\mathcal{A}X)_i = \sum_{j=1}^n A_{ij} \bullet X_j \ , \quad (i = 1, \ldots, m) \ .$$

Denote the adjoint of $\mathcal{A}$ by $\mathcal{A}^*$. Define the block diagonal matrices $X = \mathrm{Diag}(X_i)$, $Z = \mathrm{Diag}(Z_i)$ and $C = \mathrm{Diag}(C_i)$. Then $(X^\mu, \mathbf{y}^\mu, Z^\mu)$ is a solution to (3)-(4) iff it satisfies the primal-dual feasibility constraints as well as the centering condition, i.e., the system of equations below.

$$\mathcal{A}X = \mathbf{b} \tag{5}$$

$$\mathcal{A}^* \mathbf{y} + Z = C \tag{6}$$

$$X_i^\mu Z_i^\mu = \mu I \ (i \notin \mathrm{LD}) \ , \quad X_i Z_i = \hat{c}_i I \ (i \in \mathrm{LD}) \ . \tag{7}$$

The feasible constraints (5)-(6) imply that

$$C \bullet X = \sum_{i=1}^n \left(\sum_{k=1}^m y_k A_{ki} + Z_i \right) \bullet X_i \ .$$

Therefore,

$$C \bullet X^\mu - \mathbf{b}^T \mathbf{y}^\mu = X^\mu \bullet Z^\mu = \mu \sum_{i \notin \mathrm{LD}} N_i + \sum_{i \in \mathrm{LD}} \hat{c}_i N_i \ . \tag{8}$$

Let $\delta(\cdot|\,\mathrm{LD})$ denote the function

$$\delta(i|\,\mathrm{LD}) = \begin{cases} 1 & i \in \mathrm{LD} \ , \\ 0 & i \notin \mathrm{LD} \ . \end{cases}$$

For any $a \in \mathbb{R}$, we define $\nu_i(a) = \delta(i|\,\mathrm{LD})\hat{c}_i + [1 - \delta(i|\,\mathrm{LD})]a$, i.e.,

$$\nu_i(a) = \begin{cases} \hat{c}_i & i \in \mathrm{LD} \ , \\ a & i \notin \mathrm{LD} \ . \end{cases}$$

For a block diagonal matrix $M = \mathrm{Diag}(M_i)$, we define

$$\nu(a)M \stackrel{\text{def}}{=} \mathrm{Diag}\left[\nu_i(a) M_i\right] \ .$$

Next, we give existence and uniqueness of the optimal solution to the parameterized maxdet-problem. Our proof is an extension of [6, Theorem 2.1, 2.4].

Lemma 1 *Suppose that the primal-dual maxdet-problem* (1)-(2) *has an interior feasible solution. Then for all* $\mu > 0$, *the perturbed problem* (3)-(4) *has a unique solution.*

Proof: We first prove existence. Suppose $(X^0, \mathbf{y}^0, Z^0)$ is an interior feasible solution to (1)-(2). Since $(X^0, \mathbf{y}^0, Z^0)$ is feasible, using the same argument as that for (8), we get that the solution set to (3) are the minimizers of $Z^0 \bullet X - \sum_{i=1}^n \nu_i(\mu) \ln \operatorname{Det} X_i^0$ over the following set

$$L \stackrel{\text{def}}{=} \{X \succ 0\colon \mathcal{A}X = \mathbf{b},\ Z^0 \bullet X - \sum_{i=1}^n \nu_i(\mu) \ln \operatorname{Det} X_i \leq Z^0 \bullet X^0 - \sum_{i=1}^n \nu_i(\mu) \ln \operatorname{Det} X_i^0\ .\}$$

Next we give bounds on $\lambda(X)$.

$$\begin{aligned} Z^0 \bullet X^0 - \sum_{i=1}^n \nu_i(\mu) \ln \operatorname{Det} X_i^0 &\geq \lambda_{\min}(Z^0) I \bullet X - \sum_{i=1}^n \nu_i(\mu) \ln \operatorname{Det} X_i \\ &= \sum_{i=1}^n \sum_{j=1}^{N_i} \left[\lambda_{\min}(Z^0)\lambda_j(X_i) - \nu_i(\mu) \ln \lambda_j(X_i)\right]\ . \end{aligned}$$

The last term is a sum of the functions $\psi(\lambda)$ in the form $\psi(\lambda) = u\lambda - v \ln \lambda$. Note that for all $u, v > 0$, $\psi(\lambda)$ is strictly convex, tends to $+\infty$ as λ tends to 0 or $+\infty$. The inequality above shows that the last term is bounded from above; hence each eigenvalue of X is bounded. Therefore, L is a subset of a compact set of positive definite matrices. Since L is defined by linear equations and an inequality on a function that is continuous on such sets, L is compact, and $Z^0 \bullet X - \sum_{i=1}^n \nu_i(\mu) \ln \operatorname{Det} X_i$ is continuous on it. Therefore, (3) has an optimal solution X^μ. The existence of an optimal solution to (4) follows from the equations (6)-(7).

Next, we show the uniqueness. The uniqueness of X^μ and Z^μ follows from the strict convexity of the objective functions of (3)-(4). From (6) and the assumption that the null space of $\mathcal{A}^*$ is $\{0\}$, we get the uniqueness of $\mathbf{y}^0$. ■

Next, we consider the optimal set of the maxdet-problem.

Lemma 2 *Suppose that* (1)-(2) *has an interior feasible solution. Then both* (1) *and* (2) *have nonempty and bounded optimal solution sets.*

Proof: Assume $(X^0, \mathbf{y}^0, Z^0)$ is an interior feasible solution to (1)-(2). As the arguments in the preceding lemma, the optima set is not changed if we replace the objective function of (1) by $Z^0 \bullet X - \sum_{i \in \text{LD}} \hat{c}_i \ln \operatorname{Det} X_i$ and add the constraint $Z^0 \bullet X - \sum_{i \in \text{LD}} \hat{c}_i \ln \operatorname{Det} X_i \leq Z^0 \bullet X^0 - \sum_{i \in \text{LD}} \hat{c}_i \ln \operatorname{Det} X_i^0$. Consider the inequality below.

$$\begin{aligned} Z^0 \bullet X^0 - \sum_{i \in \text{LD}} \hat{c}_i \ln \operatorname{Det} X_i^0 &\geq \lambda_{\min}(Z^0) I \bullet X - \sum_{i \in \text{LD}} \hat{c}_i \ln \operatorname{Det} X_i \\ &= \sum_{i \in \text{LD}} \sum_{j=1}^{N_i} \left[\lambda_{\min}(Z^0)\lambda_j(X_i) - \hat{c}_i \ln \lambda_j(X_i)\right] + \sum_{i \notin \text{LD}} \sum_{j=1}^{N_i} \lambda_{\min}(Z^0)\lambda_j(X_i)\ . \end{aligned}$$

The right-hand-side of the last equality tends to $+\infty$ as $\lambda_j(X_i)$ tends to $+\infty$ for all $i = 1, \ldots, n$ and $j = 1, \ldots N_i$. Hence all optimal solutions of (1) lie in a compact set. Since $Z^0 X - \sum_{i \in \text{LD}} \hat{c}_i \ln \operatorname{Det} X_i$ is continuous over this compact set, we conclude that the set of optimal solutions to (1) is nonempty and bounded. Similarly, the set of optimal solutions to (2) is nonempty and bounded. ∎

Assume both (1) and (2) have a strictly feasible solution $(X^0, \mathbf{y}^0, Z^0)$, i.e., $X^0 \succ 0$, $Z^0 \succ 0$, and $(X^0, \mathbf{y}^0, Z^0)$ satisfies the feasibility constraints (5)-(6), also suppose that the null space of $\mathcal{A}^*$ is $\{0\}$; then $(X^\mu, \mathbf{y}^\mu, Z^\mu)$ converges to an optimal solution to the maxdet-problem (see [4]).

3.2 The Newton System

In this part, we derive the Newton system for the symmetric central path conditions and give its solution. To simplify notations, we use symmetric Kronecker product [2].

Let nvec denote the mapping from $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{n^2}$:

$$\text{nvec}(K) = [K_{11}, K_{12}, \dots, K_{1n}, K_{21}, \dots, K_{2n}, \dots, K_{nn}]^T \ .$$

Let svec represent the mapping from $\mathcal{S}^n$ to $\mathbb{R}^{\frac{n(n+1)}{2}}$:

$$\text{svec}(K) = \left[K_{11}, \sqrt{2}K_{12}, \dots, \sqrt{2}K_{1n}, K_{22}, \dots, \sqrt{2}K_{2n}, \dots, K_{nn}\right]^T \ .$$

We use $M \otimes N$ to represent the Kronecker product of matrices M and N. For $M, N \in \mathbb{R}^{n\times n}$, the symmetric product $M \circledast N$ is defined as a linear operator on $\mathcal{S}^n$:

$$(M \circledast N)\,\text{svec}(K) = \text{svec}\left[\frac{1}{2}(NKM^T + MKN^T)\right] \ .$$

For $P_i \in \mathbb{R}^{N_i\times N_i}$ nonsingular, let $\mathcal{H}_{P_i}$ denote the linear transformations on the *ith* block:

$$\mathcal{H}_{P_i}(M) \stackrel{\text{def}}{=} \frac{1}{2}\left[P_i M P_i^{-1} + P_i^{-T} M^T P_i^T\right] \text{, i.e.,}$$
$$\text{nvec}\left[\mathcal{H}_{P_i}(M)\right] = \frac{1}{2}\left(P_i^{-T} \otimes P_i\right)\text{nvec}(M) + \frac{1}{2}\left(P_i \otimes P_i^{-T}\right)\text{nvec}(M^T) \ .$$

Denote $P = \text{Diag}(P_i)$ $(i = 1, \dots, n)$. Correspondingly, we use $\mathcal{H}_P$ to represent the direct product of $\mathcal{H}_{P_i}$.

The Newton's method applied to the central path conditions results in

$$\begin{aligned} \mathcal{A}\Delta X =& \mathbf{b} - \mathcal{A}X \\ \mathcal{A}^*\Delta\mathbf{y} + \Delta Z =& C - \mathcal{A}^*\mathbf{y} - Z \\ X\Delta Z + Z\Delta X =& (\nu(\mu)I - XZ) \ . \end{aligned} \tag{9}$$

Because XZ is not symmetric in general, the domain and range of the function defined by the right hand side of (9) are not the same. Therefore, Newton's method is not directly applicable.

We then apply $\mathcal{H}_P$ to symmetrize the centering condition

$$\mathcal{A}\Delta X = \mathbf{r}_p \stackrel{\text{def}}{=} \mathbf{b} - \mathcal{A}X \tag{10a}$$

$$\mathcal{A}^*\Delta\mathbf{y} + \Delta Z = R_d \stackrel{\text{def}}{=} C - \mathcal{A}^*\mathbf{y} - Z \tag{10b}$$

$$\mathcal{H}_P(X\Delta Z + Z\Delta X) = R_c \stackrel{\text{def}}{=} \nu(\mu)I - \mathcal{H}_P(XZ) \ . \tag{10c}$$

Equations (10c) can also be represented as

$$\left(PZ \circledast P^{-T}\right) \operatorname{svec} \Delta X + \left(PX \circledast P^{-T}\right) \operatorname{svec} \Delta Z = \operatorname{svec}\left(\nu(\mu) I\right) - PX \circledast P^{-T} \operatorname{svec}(Z) \ . \quad (11)$$

Define the linear operator $\mathcal{E},\ \mathcal{F} : \mathcal{S}^{N_1} \times \cdots \times \mathcal{S}^{N_n} \to \mathcal{S}^{N_1} \times \cdots \times \mathcal{S}^{N_n}$ as the direct products of $\mathcal{E}_i$ and $\mathcal{F}_i$:

$$\mathcal{E}_i(M) = \frac{1}{2}\left(P_i M Z_i P_i^{-1} + P_i^{-T} Z_i M P_i^T\right) \ , \ \mathcal{F}_i(M) = \frac{1}{2}\left(P_i X_i M P_i^{-1} + P_i^{-T} M X_i P_i^T\right) \ .$$

Assume $\mathcal{E}^{-1}\mathcal{F}$ is positive definite and $\mathcal{A}_{\cdot,i}$'s are linearly independent. Then the unique solution to (10) is

$$\begin{aligned} \Delta \mathbf{y} &= \left(\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*\right)^{-1}\left[\mathbf{r}_p - \mathcal{A}\mathcal{E}^{-1}\left(R_c - \mathcal{F}R_d\right)\right] \\ \Delta Z &= R_d - \mathcal{A}^*\Delta \mathbf{y} \\ \Delta X &= \mathcal{E}^{-1}(R_c - \mathcal{F}\Delta Z) \ . \end{aligned}$$

In the following, we call P a commutative scaling matrix if PXP^T and $P^{-T}ZP^{-1}$ commute. $\mathcal{E}^{-1}\mathcal{F}$ is positive definite if P in (10) is a commutative scaling matrix. Therefore the search direction is well-defined for any interior feasible solution if P is a commutative scaling matrix in (10) (see [9, Proposition 3.1]).

3.3 The Neighborhoods of the Central Path and the Algorithm

In this part, we define our biased neighborhoods of the central path and then describe our adaptive-set path-following interior-point algorithm for the maxdet problem. Our algorithm framework includes the long-step, short-step, and Mizuno-Todd-Ye type predictor-corrector methods. Our methods require the scaling matrix to be commutative; the NT, the primal and the dual HRVW/KSH/M directions are in this family.

For $X, Z \succ 0$, we define

$$\mu(X,Z) \stackrel{\text{def}}{=} \frac{1}{\sum_{i \notin \text{LD}} N_i}\left(X \bullet Z - \sum_{i \in \text{LD}} \hat{c}_i N_i\right) \ .$$

Similarly, we denote $\nu(X,Z)I = \operatorname{Diag}\left[\nu_i(X,Z)I\right]$ with

$$\nu_i(X,Z)I = \begin{cases} \hat{c}_i I & i \in \text{LD} \ , \\ \mu(X,Z) & i \notin \text{LD} \ . \end{cases}$$

Note that our definition of $\mu(X,Z)$ is different from that in [9]; however, the values are the same when X and Z satisfy the centering condition (7).

The biased neighborhoods. In the complexity analysis for SDP, iterates are restricted to a neighborhood of the central path. The distance of the iterate from the central path is usually measured by the Frobenius norm, the infinity norm, or the semi-infinity norm. The semi-infinity norm $\|M\|_{-\infty}$ is defined to be the negative of the smallest eigenvalue of M. This measurement is called semi-infinity norm, although it is not a semi-norm at all. The neighborhood associated with the Frobenius norm is the narrowest and that associated with the semi-infinity norm is the widest. Note that when $X \succeq 0$, $X^{1/2}ZX^{1/2}$ and XZ are similar. Since the former is a symmetric matrix, all its eigenvalues are real. It follows that all the eigenvalues of XZ are real. We define the narrow and wide neighborhoods with respect to given constants γ and μ below.

$$\mathcal{N}_{F(\gamma)}(\mu) \stackrel{\text{def}}{=} \{(X, \mathbf{y}, Z) \colon X \succ 0,\ Z \succ 0,\ AX = \mathbf{b},\ A^*\mathbf{y} + Z = C, \\ \left\| X^{1/2} Z X^{1/2} - \nu(\mu) \right\|_{\mathrm{F}} \leq \gamma\mu\ .\}$$

$$\mathcal{N}_{-\infty(\gamma)}(\mu) \stackrel{\text{def}}{=} \{(X, \mathbf{y}, Z) \colon X \succ 0,\ Z \succ 0,\ AX = \mathbf{b},\ A^*\mathbf{y} + Z = C, \\ \max_{1 \leq i \leq n} \left[\nu_i(\mu) - \lambda_{\min}(X_i^{1/2} Z_i X_i^{1/2}) \right] \leq \gamma\mu\ .\}$$

Note that $\max_{1 \leq i \leq n} \left[\nu_i(\mu) - \lambda_{\min}(X_i^{1/2} Z_i X_i^{1/2}) \right] \leq \gamma\mu$ is equivalent to: for all $i = 1, \ldots, n$, $\lambda_{\min}(X_i Z_i) \geq \nu_i(\mu) - \gamma\mu$.

It is easy to see that $\max_{1 \leq i \leq n} \left[\nu_i(\mu) - \lambda_{\min}(X_i^{1/2} Z_i X_i^{1/2}) \right] \leq \left\| X^{1/2} Z X^{1/2} - \nu(\mu) \right\|_{\mathrm{F}}$. Hence the name wide and narrow neighborhoods.

Next, we describe our algorithm for the max-det problem.

The adaptive-set path-following algorithm. Given an initial feasible interior point $(X^0, \mathbf{y}^0, Z^0)$ and accuracy threshold $\epsilon > 0$, our algorithm finds an ϵ-optimal solution $(\tilde{X}, \tilde{\mathbf{y}}, \tilde{Z})$ in the sense that $\mu(\tilde{X}, \tilde{Z}) \leq \epsilon$, $\mathcal{A}\tilde{X} = \mathbf{b}$, $\mathcal{A}^*\tilde{\mathbf{y}} + \tilde{Z} = 0$.

Initialization Set $d = 0$. Choose LD^d and μ^0 so that $\mathrm{LD}^d = \{i \colon i \in \mathrm{LD},\ \hat{c}_i > \mu^0\}$, $\mu_0 = \frac{1}{\sum_{i \notin \mathrm{LD}^d}} \left(X^0 \bullet Z^0 - \sum_{i \in \mathrm{LD}^d} \hat{c}_i N_i \right)$ [1]

Outer-loop

1. While $\mathrm{LD} \neq \mathrm{LD}^d$, do the following.
   (a) Set $\tilde{c}^d = \max(\hat{c}_i \in \mathrm{LD} \colon i \notin \mathrm{LD}^d)$.
   (b) **Inner-loop** Apply the long-step, or short-step, or predictor-corrector method for the max-det problem, which is described below, with the determinant set being LD^d, to reduce $\mu(X, Z)$ to $\tilde{c}^d$.
   (c) Update $\mathrm{LD}^d \longleftarrow \mathrm{LD}^d \cup \{i \colon \hat{c}_i = \tilde{c}^d,\ i \in \mathrm{LD}\}$.
   (d) $d \longleftarrow d + 1$.
2. When $\mathrm{LD} = \mathrm{LD}^d$, reduce the duality gap $\mu(X, Z)$ to ϵ or lower with the long-step, or short-step, or predictor-corrector method.

Below are the inner-loops for the above algorithm framework. For simplicity, we don't differentiate $\tilde{c}^d$ with ϵ, LD^d with LD, in our algorithm description.

The long-step method. Set $k = 0$, $\mu^k = \frac{1}{\sum_{i \notin \mathrm{LD}} N_i} \left(X^0 \bullet Z^0 - \sum_{i \in \mathrm{LD}} \hat{c}_i N_i \right)$. Choose parameters $0 < \sigma < 1$, $0 < \gamma < 1$. Assume $(X^0, \mathbf{y}^0, Z^0) \in \mathcal{N}_{-\infty(\gamma)}(\mu^0)$, and $\hat{c}_i \geq \mu^0$ $(i \in \mathrm{LD})$.

Do while $\mu^k > \epsilon$.

1. Choose a commutative scaling matrix P^k and solve (10) with $\mu = \sigma\mu^k$.

[1] If $\mathrm{LD}^d = \{1, \ldots, n\}$, we either set $\epsilon = \min_{i=1,\ldots,n} \hat{c}_i$, or add a dummy variable X_{n+1}.

2. Find step length α^k so that the iterate

$$(X^{k+1}, \mathbf{y}^{k+1}, Z^{k+1}) = (X^k, \mathbf{y}^k, Z^k) + \alpha^k(\Delta X, \Delta \mathbf{y}, \Delta Z) \ ,$$
$$\mu^{k+1} = \frac{1}{\sum_{i \notin \mathrm{LD}} N_i} \left(X^{k+1} \bullet Z^{k+1} - \sum_{i \in \mathrm{LD}} \hat{c}_i N_i \right) \ ,$$

remains in $\mathcal{N}_{-\infty(\gamma)}(\mu^{k+1})$.

3. Set $k \longleftarrow k+1$.

Note that the search direction is always well-defined due to our choice of the scaling matrices. The choice of μ^k in the algorithm implies that for any $(X^k, \mathbf{y}^k, Z^k) \in \mathcal{N}_{-\infty(\gamma)}(\mu^k)$, and for any $i = 1, \ldots, n$:

$$-\left(\sum_{i=1}^{n} N_i - 1 \right) \gamma \mu^k \leq \nu_i(\mu^k) - \lambda_j(X_i^k Z_i^k) \leq \gamma \mu^k \quad (j = 1, \ldots, N_i) \ . \tag{12}$$

Therefore, as $\mu^k \to 0$, $(X^k, \mathbf{y}^k, Z^k)$ approximates an optimal solution of (1)-(2).

The short-step method. For all $k \geq 0$, the scaling matrix P^k is chosen to be in the commutative class, i.e., PXP^T and $P^{-T}ZP^{-1}$ commute. The iterates $(X^k, \mathbf{y}^k, Z^k)$ is kept in $\mathcal{N}_{F(\gamma)}(\mu)$. The step-size $\alpha^k = 1$. When $\mathrm{LD}^d \neq \mathrm{LD}$, the step-size is chosen as below.

$$\alpha^k = \begin{cases} \frac{\mu^k - \tilde{c}^d}{\mu^k} & \tilde{c}^d > \mu^k \ ; \\ 1 & \tilde{c}^d \leq \mu^k \ . \end{cases}$$

The parameters σ and γ is specified in the next section to ensure polynomial-time complexity of the algorithm.

The predictor-corrector method. Choose constant $\tau \in (0, 0.25)$. Assume $(X^0, \mathbf{y}^0, Z^0) \in \mathcal{N}_{F(\tau)}(\mu^0)$. For every $k \geq 0$ even, let $\sigma = 0$. Choose the stepsize α^k to be the largest $\tilde{\alpha} > 0$ such that $(X^k, \mathbf{y}^k, Z^k) + \alpha(\Delta X, \Delta \mathbf{y}, \Delta Z) \in \mathcal{N}_{F(2\tau)}(\mu)$ for all $\alpha \in [0, \tilde{\alpha}]$. When $\mathrm{LD}^d \neq \mathrm{LD}$, the step-size is chosen as the following.

$$\alpha^k = \begin{cases} \frac{\mu^k - \tilde{c}^d}{(1-\sigma)\mu^k} & \tilde{c}^d > \sigma \mu^k \ ; \\ \tilde{\alpha} & \tilde{c}^d \leq \sigma \mu^k \ . \end{cases}$$

For every $k > 0$ odd, let $\sigma = 1$ and $\alpha^k = 1$.

4 Complexity of the Algorithm

In this part, we summarize the complexity of our long-step, short-step, and predictor-corrector interior-point path-following algorithm for the maxdet problem.

4.1 The Long Step Algorithm

Let $G_\infty \stackrel{\text{def}}{=} \sup\{\text{cond}_{\max}(P^k X^k P^{k^T} P^k Z^{k^{-1}} P^{k^T})\colon k = 0, 1, 2, \ldots\}$. Assume in addition, G_∞ is upper bounded for all the iterates. Suppose each P^k is chosen such that PXP^T and $PZ^{-1}P^T$ commute. Then we can choose the step-size $\alpha^k \geq \alpha^*$, where α^* is defined as

$$\alpha^* \stackrel{\text{def}}{=} \min\left\{1,\ \frac{\sigma\gamma}{\sqrt{G_\infty}}\left[\left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right)\sum_{i \notin \text{LD}} N_i + \sum_{i \in \text{LD}} \frac{\gamma \hat{c}_i N_i}{\hat{c}_i - \gamma\mu_0}\right]^{-1}\right\}.$$

In addition,

$$\mu^{k+1} \stackrel{\text{def}}{=} \mu\left(X^{k+1}, Z^{k+1}\right) = [1 - (1-\sigma)\alpha^k]\mu^k\ .$$

Therefore, the long-step method terminates in at most $\mathcal{O}[\sqrt{G_\infty} N \ln(\mu^0/\epsilon)]$ iterations. Especially, the complexity bounds on the long-step path-following algorithm for the maxdet problem for the NT direction, the HRVW/KSH/M diction, and the dual HRVW/KSH/M direction are $\mathcal{O}[N \ln(\mu^0/\epsilon)]$, $\mathcal{O}[N^{3/2} \ln(\mu^0/\epsilon)]$, and $\mathcal{O}[N^{3/2} \ln(\mu^0/\epsilon)]$, respectively.

4.2 The Short-Step Method

Let $0 < \gamma < 1$ and $0 < \sigma < 1$ be constants satisfying

$$\left[\gamma + (1-\sigma)\sqrt{\sum_{i \notin \text{LD}} N_i}\right]^2 \leq 2(1-\gamma)\sigma\gamma. \tag{13}$$

Assume that $(X, \mathbf{y}, S) \in \mathcal{N}_{F(\gamma)}(\mu)$ and $(\Delta X, \Delta\mathbf{y}, \Delta S)$ is the solution to (10) with μ on its right-hand-side replaced by $\sigma\mu$. Suppose that the scaling matrix P is a commutative scaling matrix. Then

1. $\mu\left(X(1), Z(1)\right) \stackrel{\text{def}}{=} \mu\left(X + \Delta X, Z + \Delta Z\right) = \sigma\mu\left(X, Z\right)$;
2. $(X(1), \mathbf{y}(1), Z(1)) \stackrel{\text{def}}{=} (X + \Delta X, \mathbf{y} + \Delta\mathbf{y}, Z + \Delta Z) \in \mathcal{N}_{F(\gamma)}(\mu(1))$.

Therefore, suppose that the cardinality of the set LD is no more than $\mathcal{O}\left(\sqrt{N}\right)$. Then the short-step method terminates in at most $\mathcal{O}\left[\sqrt{N}\ln(\mu^0/\epsilon)\right]$ steps.

4.3 The Predictor-Corrector Method

1. For a constant $0 < \tau < 1/2$, suppose that $(X, \mathbf{y}, Z) \in \mathcal{N}_{F(\tau)}\left(\mu(X, Z)\right)$, and $(\Delta X, \Delta\mathbf{y}, \Delta S)$ is the solution to (10) with μ on its right-hand-side replaced by 0. Suppose that the scaling matrix P is a commutative scaling matrix. Let $\tilde{\alpha}$ denote the unique positive root of $\alpha^2 \frac{\left[\tau + \sqrt{\sum_{i \notin \text{LD}} N_i}\right]^2}{2(1-\tau)} + \alpha\tau - \tau$. Then, for any $\alpha \in [0, \tilde{\alpha}]$, we have

   (a) $\mu(\alpha) \stackrel{\text{def}}{=} \mu\left(X(\alpha), Z(\alpha)\right) = (1-\alpha)\mu\left(X, Z\right)$.
   (b) $[X(\alpha), \mathbf{y}(\alpha), Z(\alpha)] \in \mathcal{N}_{F(2\tau)}(\mu(\alpha))$.

   In addition, $\tilde{\alpha} = 1/\mathcal{O}(N^{1/2})$.

2. Suppose that $(X, \mathbf{y}, Z) \in \mathcal{N}_{F(2\tau)}(\mu)$ for some constant $0 < \tau < 0.25$. Let $\sigma = 1$. Assume that $(\Delta X, \Delta \mathbf{y}, \Delta S)$ is the solution to (10). Suppose that the scaling matrix P is a commutative scaling matrix. Then

   (a) $\mu(1) \stackrel{\text{def}}{=} \mu(X + \Delta X, Z + \Delta Z) = \mu(X, Z)$;
   (b) $(X(1), \mathbf{y}(1), Z(1)) \in \mathcal{N}_{F(\tau)}(\mu(1))$.

Hence, suppose that the cardinality of LD is less than $\mathcal{O}\left(\sqrt{N}\right)$. Then the predictor-corrector method terminates in at most $\mathcal{O}\left[\sqrt{N}\ln(\mu^0/\epsilon)\right]$ steps.

5 Copyright

Acknowledgment

I thank a fellowship for foreign researchers from JSPS and Grant-in-Aid for JSPS Fellows (16-04291). I thank Prof. Takashi Tsuchiya for motivation and discussion. I thank Dr. Mituhiro Fukuda for pointing out the constant in the dual, Prof. Chek Beng Chua for pointing out the case $\mathrm{LD}^d = \{1, \ldots, n\}$.

References

[1] Farid Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.*, 5(1):13–51, 1995.

[2] Farid Alizadeh, Jean-Pierre A. Haeberly, and Michael L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM J. Optim.*, 8(3):746–768 (electronic), 1998.

[3] Tadayoshi Fushiki, Shingo Horiuchi, and Takashi Tsuchiya. A new computational approach to density estimation with semidefinite programming. Research Memorandum 898, The Institute of Statistical Mathematics, Tokyo, Japan, November 2003.

[4] L. M. Graña Drummond and Y. Peterzil. The central path in smooth convex semidefinite programs. *Optimization*, 51(2):207–233, 2002.

[5] Trevor Hastie, Robert Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning*. Springer, July 2001.

[6] Renato Monteiro and Michael Todd. Path-following methods. In *Handbook of semidefinite programming*, volume 27 of *Internat. Ser. Oper. Res. Management Sci.*, pages 267–306. Kluwer Acad. Publ., Boston, MA, 2000.

[7] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[8] Claude E. Shannon and Warren Weaver. *A Mathematical Theory of Communication*. University of Illinois Press, Champaign, IL, USA, 1963.

[9] Kim-Chuan Toh. Primal-dual path-following algorithms for determinant maximization problems with linear matrix inequalities. *Comput. Optim. Appl.*, 14(3):309–330, 1999.

[10] Takashi Tsuchiya and Yu Xia. An extension of the standard polynomial-time primal-dual path-following algorithm to the weighted determinant maximization problem with semidefinite constraints. Technical Report Research Memorandum No. 980, The Institute of Statistical Mathematics, Japan, 2006. To appear in *Pacific Journal of Optimization*.

[11] Lieven Vandenberghe, Stephen Boyd, and Shao-Po Wu. Determinant maximization with linear matrix inequality constraints. *SIAM J. Matrix Anal. Appl.*, 19(2):499–533 (electronic), 1998.