Uncorrelated Gardner's Problem

Ruilin Li

1 Introduction

Consider the following problem: let $\mathbf{g^1}, \ldots, \mathbf{g^m} \in \mathbb{R}^n$ be i.i.d. standard n-dimensional Gaussian random variables, and $\kappa \in \mathbb{R}$ is a fixed number. Define

$$A = \bigcap_{l=1}^{m} \{\mathbf{x} \in S^{n-1} : \mathbf{g}^l \cdot \mathbf{x} \geq \kappa\}$$

where S^{n-1} is the $n-1$ dimensional sphere embedded in $\mathbb{R}^n$. Clearly with probability one the set A will shrink as m increases. The problem of interest is that, if we set $m = \alpha n$, how large can α be such that A is non-empty with high probability. By high probability, we mean that there exists a constant $c > 0$ such that

$$\mathbb{P}(A \text{ is empty}) < e^{-cn}. \tag{1}$$

Such problem occurs in the perceptron model, which is defined by the following dynamics. Suppose for $i = 1, \ldots, n$, $H_i^t \in \{-1, 1\}$ are the states of the neuron at time t, and for each $1 \leq i, j \leq n$, $x_j^i \in \mathbb{R}$ is the interaction strength from neuron j to neuron i. We require that for each i, $x_i^i = 0$, and

$$\sum_{j=1}^{n} (x_j^i)^2 = 1$$

The ith neuron fires at time t if $H_i^t = 1$, and does not fire if $H_i^t = -1$. The state of a neuron is updated according to:

$$H_i^{t+1} = \text{sign}\Big(\sum_{j=1, j \neq i}^{n} H_j^t x_j^i\Big).$$

Given interaction strength x^i_j, the pattern $\mathbf{H} = (H_1, \ldots, H_n)$ is memorized by the perceptron if $\mathbf{H}$ is a fixed point of the above dynamics, that is for each $i = 1, \ldots, n$

$$H_i^{t+1}\Big(\sum_{j=1}^{n} H_j^t x_j^i\Big) \geq 0.$$

Here to ensure stability we also require

$$H_i\Big(\sum_{j=1, j\neq i}^{n} H_j x_j^i\Big) \geq \kappa \tag{2}$$

for a fixed constant $\kappa > 0$. We are interested in the generic capacity of the perceptron: how many random patterns can we take, so that there is a high probability that there exists a set of interaction strength satisfying (2) for all of the random patterns. Although in our definition the states of the neurons can only take -1 and 1, here we will relax this assumption and allow the states to be any real number. Moreover, we will take the random pattern $\mathbf{g} \in \mathbb{R}^n$ to come from the n-dimensional standard Gaussian distribution.

To be more precise, let $m = \alpha n$ to be the number of random patterns. For $l = 1, \ldots, m$, $\mathbf{g}^l \in \mathbb{R}^n$ is a random pattern generated from a standard Gaussian distribution. Let $\mathbf{g}^l_i \in \mathbb{R}$ be the ith coordinate of $\mathbf{g}^l$, and $\mathbf{g}^l_{-i} \in \mathbb{R}^{n-1}$ be the rest of the coordinates of $\mathbf{g}^l$. We are then interested in the probability of the event:

$$E = \{\forall i, \exists \mathbf{x} \in S^{n-2} \text{ such that } \operatorname{sign}(\mathbf{g}^l_i)(\mathbf{g}^l_{-i}\cdot\mathbf{x}) \geq \kappa \;\forall l\}$$

which means that the perceptron can successfully remember all of the m random patterns. Then the negation of E can be written as:

$$E^c = \{\exists i \text{ such that } \forall \mathbf{x} \in S^{n-2},\ \operatorname{sign}(\mathbf{g}^l_i)(\mathbf{g}^l_{-i}\cdot\mathbf{x}) < \kappa \text{ for some } l\}.$$

Therefore

$$\begin{aligned}
\mathbb{P}(E^c) &\leq \sum_{i=1}^{n} \mathbb{P}(\forall\, \mathbf{x} \in S^{n-2}, \exists l \leq m \text{ such that } \operatorname{sign}(\mathbf{g}^l_i)(\mathbf{g}^l_{-i}\cdot\mathbf{x}) < \kappa) \\
&= n\mathbb{P}(\forall\, \mathbf{x} \in S^{n-2}, \exists l \leq m \text{ such that } \operatorname{sign}(\mathbf{g}^l_1)(\mathbf{g}^l_{-1}\cdot\mathbf{x}) < \kappa) \\
&= n\mathbb{P}(\forall\, \mathbf{x} \in S^{n-2}, \exists l \leq m \text{ such that } (\mathbf{g}^l_{-1}\cdot\mathbf{x}) < \kappa) \\
&= n\mathbb{P}\Big(\bigcap_{l=1}^{m}\{\mathbf{x} \in S^{n-2} : \mathbf{g}^l_{-1}\cdot\mathbf{x} \geq \kappa\} \text{ is empty}\Big).
\end{aligned}$$

Similarly

$$\mathbb{P}(E) \leq \mathbb{P}\Big(\bigcap_{l=1}^{m}\{\mathbf{x} \in S^{n-2} : \mathbf{g}^l_{-1}\cdot\mathbf{x} \geq \kappa\} \text{ is not empty}\Big).$$

It is clear now that the perceptron capacity problem above is exactly the same problem at the beginning of the section, except $n-1$ in (1) becomes $n-2$ here. This change will not affect our analysis on α. If for some particular choice of α I can show that (1) is true, then $\mathbb{P}(E^c)$ also decays in exponentially in n (since multiplying by n does not change exponential decay rate). Therefore, the rest of the write-up will focus on solving the α so that (1) is satisfied. I mentioned an upper bound for $P(E)$ because more can be said about (1). In fact, as you will see in the next section, one of $\mathbb{P}(\text{A is empty})$ and $\mathbb{P}(\text{A is not empty})$ will have exponential decay.

2 Solution To the Uncorrelated Gardner's Problem

We begin this section by stating the solution to the problem in the first section.

Theorem 1. *Let $A,m,n,$ and α be defined same as in section 1. Let g be a standard one dimensional Gaussian random variable and $(g+\kappa)_+$ be the positive part of the random variable $g+\kappa$. Then,*

$$\alpha\mathbb{E}(g+\kappa)_+^2 > 1, \kappa \in \mathbb{R} \implies \exists c > 0 \text{ such that } \mathbb{P}(A \text{ is empty}) \geq 1 - e^{-cn} \tag{3}$$

and

$$\alpha\mathbb{E}(g+\kappa)_+^2 < 1, \kappa > 0 \implies \exists c > 0 \text{ such that } \mathbb{P}(A \text{ is not empty}) \geq 1 - e^{-cn} \tag{4}$$

2.1 Proof of the First Part of Theorem 1

Assume $\alpha\mathbb{E}(g+\kappa)_+^2 > 1$, and $\kappa \in \mathbb{R}$. Let G be a $m \times n$ matrix whose entries are i.i.d. standard Gaussian random variables, and $\mathbf{1} = (1,\ldots,1)^T \in \mathbf{R}^m$. Then A is non-empty if and only if there exists $\mathbf{x} \in S^{n-1}$ such that

$$G\mathbf{x} \geq \kappa\mathbf{1}$$

where the inequality of two vectors should be interpreted as coordinate-wise inequalities. We notice that the above condition is equivalent to the following one, let $\lambda \in \mathbf{R}^m$

$$\min_{\mathbf{x}} \max_{\lambda \geq 0} \lambda^T(\kappa\mathbf{1} - G\mathbf{x}) \leq 0, \text{with } \|\mathbf{x}\| = 1, \|\lambda\| \leq 1$$

For the reminder of this section, I will assume that $\|\mathbf{x}\| = 1, \|\lambda\| \leq 1$. Therefore, to show (3) it suffices to show that there exists $\delta, c > 0$, such that

$$\mathbb{P}\Big(\min_{\mathbf{x}} \max_{\lambda \geq 0} \lambda^T(\kappa\mathbf{1} - G\mathbf{x}) \geq \delta\Big) \geq 1 - e^{-cn}$$

We start with the following observation. Let g be a one-dimensional standard Gaussian random variable, independent of the entries in G, then for any $\delta, \varepsilon > 0$,

$$\mathbb{P}\Big(\min_{\mathbf{x}}\max_{\lambda\geq 0}\lambda^T(\kappa\mathbf{1}-G\mathbf{x})+g-\varepsilon\sqrt{n}+\delta\geq\delta\Big)\leq\mathbb{P}\Big(\min_{\mathbf{x}}\max_{\lambda\geq 0}\lambda^T(\kappa\mathbf{1}-G\mathbf{x})\geq\delta\Big)+\mathbb{P}(g\geq\varepsilon\sqrt{n}-\delta) \quad (5)$$

The second term on the right-hand side is easy to bound from above. For now we will first focus on the probability on the left-hand side. To do this, we will use the following theorem:

Theorem 2. *(Gordon's Inequality) Let $(X_{ij})_{i\leq n,j\leq m}$, $(Y_{ij})_{i\leq n,j\leq m}$ be centerd Gaussian random variables, such that*

1. $\mathbb{E}(X_{ij}^2)=\mathbb{E}(Y_{ij}^2)$ for all $i\leq n, j\leq m$
2. $\mathbb{E}(X_{ij}X_{ik})\geq\mathbb{E}(Y_{ij}Y_{ik})$ for all $i\leq n, j\leq m$
3. $\mathbb{E}(X_{ij}X_{lk})\leq\mathbb{E}(Y_{ij}Y_{lk})$ for $l\neq i$

then for any choice of $\beta_{ij}\in\mathbb{R}$

$$\mathbb{P}\Big(\min_i\max_j(X_{ij}-\beta_{ij})\geq 0\Big)\leq\mathbb{P}\Big(\min_i\max_j(Y_{ij}-\beta_{ij})\geq 0\Big)$$

Here we will use $-\lambda_i^T G\mathbf{x}_j+g$ to be the Y_{ij} in the theorem, and $\lambda_i^T\mathbf{g}+\mathbf{x}_j^T\mathbf{h}$ to be the X_{ij} in the theorem, where $\mathbf{g}\in\mathbb{R}^m,\mathbf{h}\in\mathbb{R}^n$ are independent standard Gaussian random variables. Let's check if the conditions in theorem 2 are satisfied. For each i,j,l,k, since the entries of G are i.i.d. standard Gaussian random variables, it is nor difficult to verify that

$$\mathbb{E}(-\lambda_i^T G\mathbf{x}_j+g)(-\lambda_l^T G\mathbf{x}_k+g)=(\lambda_i^T\lambda_l)(\mathbf{x}_j^T\mathbf{x}_k)+1$$

and

$$\mathbb{E}(\lambda_i^T\mathbf{g}+\mathbf{x}_j^T\mathbf{h})(\lambda_l^T\mathbf{g}+\mathbf{x}_k^T\mathbf{h})=(\lambda_i^T\lambda_l)+(\mathbf{x}_j^T\mathbf{x}_k)$$

If $i=l$, $j=k$

$$\mathbb{E}(-\lambda_i^T G\mathbf{x}_j+g)^2=2=\mathbb{E}(\lambda_i^T\mathbf{g}+\mathbf{x}_j^T\mathbf{h})^2$$

If $i=l$,

$$\mathbb{E}(-\lambda_i^T G\mathbf{x}_j+g)(-\lambda_i^T G\mathbf{x}_k+g)=1+(\mathbf{x}_j^T\mathbf{x}_k)=\mathbb{E}(\lambda_i^T\mathbf{g}+\mathbf{x}_j^T\mathbf{h})(\lambda_i^T\mathbf{g}+\mathbf{x}_k^T\mathbf{h})$$

If $i\neq l$, by Cauchy-Schwartz inequality

$$\begin{aligned}&\mathbb{E}(-\lambda_i^T G\mathbf{x}_j+g)(-\lambda_l^T G\mathbf{x}_k+g)-\mathbb{E}(\lambda_i^T\mathbf{g}+\mathbf{x}_j^T\mathbf{h})(\lambda_l^T\mathbf{g}+\mathbf{x}_k^T\mathbf{h})\\&=(\lambda_i^T\lambda_l)(\mathbf{x}_j^T\mathbf{x}_k)+1-(\mathbf{x}_j^T\mathbf{x}_k)-(\lambda_i^T\lambda_l)\\&=(1-\lambda_i^T\lambda_l)(1-\mathbf{x}_j^T\mathbf{x}_k)\geq 0\end{aligned}$$

Therefore, by Gordon's inequality, we have

$$\mathbb{P}\Big(\min_{\mathbf{x}}\max_{\lambda\geq 0}\lambda^T\kappa\mathbf{1}-\lambda^T G\mathbf{x}+g-\varepsilon\sqrt{n}\geq 0\Big)\geq\mathbb{P}\Big(\min_{\mathbf{x}}\max_{\lambda\geq 0}\lambda^T\kappa\mathbf{1}+\lambda^T\mathbf{g}+\mathbf{x}^T\mathbf{h}-\varepsilon\sqrt{n}\geq 0\Big)$$

Now we can find the optimum choice of $\mathbf{x}$ and λ on the right-hand side explicitly. The optimum is achieved when $\mathbf{x} = -\mathbf{h}/\|\mathbf{h}\|$, λ is in the same direction of $(\mathbf{g}+\kappa)_+ = (\max\{g_i+\kappa, 0\})_{i=1}^m$ (g_i is the ith coordinate of $\mathbf{g}$).

$$\min_{\mathbf{x}} \max_{\lambda\geq 0} \lambda^T\kappa\mathbf{1} + \lambda^T\mathbf{g} + \mathbf{x}^T\mathbf{h} - \varepsilon\sqrt{n} = \sqrt{\sum_{i=1}^m (g_i+\kappa)_+^2} - \sqrt{\sum_{i=1}^n h_i^2} - \varepsilon\sqrt{n}$$

By our assumption, we can choose $\varepsilon > 0$ small enough such that

$$\sqrt{\alpha\mathbb{E}(g+\kappa)_+^2} - 1 - 3\varepsilon > 0 \implies \sqrt{m\mathbb{E}(g+\kappa)_+^2} - \sqrt{n} - 2\varepsilon\sqrt{n} \geq \varepsilon\sqrt{n}$$

Therefore

$$\begin{aligned}
&\mathbb{P}\Big(\min_{\mathbf{x}} \max_{\lambda\geq 0} \lambda^T\kappa\mathbf{1} - \lambda^T G\mathbf{x} + g - \varepsilon\sqrt{n} \geq 0\Big) \\
&\geq \mathbb{P}\Big(\min_{\mathbf{x}} \max_{\lambda\geq 0} \lambda^T\kappa\mathbf{1} + \lambda^T\mathbf{g} + \mathbf{x}^T\mathbf{h} - \varepsilon\sqrt{n} \geq 0\Big) \\
&= \mathbb{P}\Big(\sqrt{\sum_{i=1}^m (g_i+\kappa)_+^2} - \sqrt{\sum_{i=1}^n h_i^2} - \varepsilon\sqrt{n} \geq 0\Big) \\
&\geq \mathbb{P}\Big(\sqrt{\sum_{i=1}^m (g_i+\kappa)_+^2} - \sqrt{\sum_{i=1}^n h_i^2} \geq \sqrt{m\mathbb{E}(g+\kappa)_+^2} - \sqrt{n} - 2\varepsilon\sqrt{n}\Big) \\
&= 1 - \mathbb{P}\Big(\sqrt{\sum_{i=1}^m (g_i+\kappa)_+^2} - \sqrt{\sum_{i=1}^n h_i^2} < \sqrt{m\mathbb{E}(g+\kappa)_+^2} - \sqrt{n} - 2\varepsilon\sqrt{n}\Big) \\
&\geq 1 - \mathbb{P}\Big(\sqrt{\sum_{i=1}^m (g_i+\kappa)_+^2} \leq \sqrt{m\mathbb{E}(g+\kappa)_+^2} - \varepsilon\sqrt{n}\Big) - \mathbb{P}\Big(\sqrt{\sum_{i=1}^n h_i^2} \geq \sqrt{n} + \varepsilon\sqrt{n}\Big)
\end{aligned} \tag{6}$$

Take the square of the two inequalities in the last line, we have

$$\sum_{i=1}^m (g_i+\kappa)_+^2 \leq m\mathbb{E}(g+\kappa)_+^2 - \Big(2\varepsilon\sqrt{\alpha\mathbb{E}(g+\kappa)_+^2} - \varepsilon^2\Big)n$$

$$\sum_{i=1}^n h_i^2 \geq n + (2\varepsilon + \varepsilon^2)n^2$$

Set $C_\varepsilon = 2\varepsilon\sqrt{\alpha\mathbb{E}(g+\kappa)_+^2} - \varepsilon^2$, $C'_\varepsilon = 2\varepsilon + \varepsilon^2$. Let $X_i = -(g_i+\kappa)_+^2$ or $X_i = h_i^2$, $N = n$ or $N = m$, and $C = C_\varepsilon$ or $C = C'_\varepsilon$. Then above showed that both probabilities in (6) can be written in the form

$$\mathbb{P}\Big(\sum_{i=1}^N (X_i - \mathbb{E}X_i) \geq Cn\Big)$$

where X_i are i.i.d. random variables. By Markov's inequality, for any $t>0$:

$$\begin{aligned}\mathbb{P}\Big(\sum_{i=1}^{N}(X_i-\mathbb{E}X_i)\geq Cn\Big) &= \mathbb{P}\Big(\exp\big(\sum_{i=1}^{N}(X_i-\mathbb{E}X_i)t\big)\geq e^{tCn}\Big)\\ &\leq e^{-tCn}\Big[\mathbb{E}\Big(\exp(t(X_1-\mathbb{E}X_1))\Big)\Big]^N\end{aligned} \tag{7}$$

Apply Taylor expansion to $\mathbb{E}\Big(\exp(t(X_1-\mathbb{E}X_1))\Big)$, the order one term vanishes we have

$$\mathbb{E}\Big(\exp(t(X_1-\mathbb{E}X_1))\Big)\leq 1+\sum_{k=2}^{\infty}\frac{t^k}{k!}\mathbb{E}|X_1-\mathbb{E}X_1|^k$$

Moreover, for each $k\geq 2$, we have

$$\begin{aligned}\mathbb{E}\big|h_1^2-1\big|^k &\leq 2^k+2^k\mathbb{E}h_1^{2k}\\ &= 2^k+2^k(2k-1)(2k-3)\cdots 1\\ &< 2^k+2^k(2k)(2k-2)(2k-4)\cdots 2\\ &= 2^k+2^{2k}k!\leq a^k k!\end{aligned}$$

for some $a>0$. By similar arguments, above holds when $X_i=-(g_i+\kappa)_+^2$ (possible with a different choice of a). Therefore,

$$\begin{aligned}\mathbb{E}\Big(\exp(t(X_1-\mathbb{E}X_1))\Big) &\leq 1+\sum_{k=2}^{\infty}\frac{t^k}{k!}\mathbb{E}|X_1-\mathbb{E}X_1|^k\\ &\leq 1+\sum_{k=2}^{\infty}\frac{t^k}{k!}a^k k! = 1+\frac{t^2a^2}{1-ta}\end{aligned}$$

Choose $0<t<1/2a$, then

$$\mathbb{E}\Big(\exp(t(X_1-\mathbb{E}X_1))\Big)\leq 1+\frac{t^2a^2}{1-ta}\leq 1+2t^2a^2\leq e^{2t^2a^2}$$

Now go back to (7), if $N=n$

$$\mathbb{P}\Big(\sum_{i=1}^{N}(X_i-\mathbb{E}X_i)\geq Cn\Big)\leq (e^{-tC+2t^2a^2})^n$$

Recall that $C\sim O(\varepsilon)$, so we can choose $\varepsilon>0$ small enough such that $t=C/4a^2$ and $t\leq 1/2a$. Then

$$\mathbb{P}\Big(\sum_{i=1}^{N}(X_i-\mathbb{E}X_i)\geq Cn\Big)\leq e^{-nC^2/8a^2}$$

Similarly, if $N = m = n\alpha$, we can further reduce $\varepsilon > 0$ (if needed), so that $t' = C/(4a^2\alpha)$, and $t' < 1/2a$. Then by (7)

$$\mathbb{P}\Big(\sum_{i=1}^{N}(X_i - \mathbb{E}X_i) \geq Cn\Big) \leq (e^{-tC+2t^2a^2\alpha})^n \leq e^{-nC^2/8\alpha a^2}$$

We have showed that both probabilities in the last line of (6) decreases exponentially in n. Moreover $\mathbb{P}(g \geq \varepsilon\sqrt{n} - \delta)$ decays exponentially in n. Therefor, by (5) there exists $c > 0$ such that

$$\mathbb{P}\Big(\min_{\mathbf{x}} \max_{\lambda \geq 0} \lambda^T(\kappa\mathbf{1} - G\mathbf{x}) \geq \delta\Big) \geq 1 - e^{-cn}$$

which implies

$$\mathbb{P}(A \text{ is empty}) \geq 1 - e^{-cn}.$$

2.2 Proof of the Second Part of Theorem 1

Now assume $\alpha\mathbb{E}(g+\kappa)_+^2 < 1$, and $\kappa \geq 0$. We want to show that the event $\{\exists\mathbf{x}$ such that $G\mathbf{x} \geq \mathbf{1}\kappa\}$ has high probability. I have shown in the last section that this event is equivalent to the event $\{\min_{\mathbf{x}} \max_{\lambda \geq 0} \lambda^T(\kappa\mathbf{1} - G\mathbf{x}) \leq 0\}$ with constraints $\|\mathbf{x}\| = 1, \|\lambda\| \leq 1$. Here this condition is also equivalent to $\{\min_{\mathbf{x}} \max_{\lambda \geq 0} \lambda^T(\kappa\mathbf{1} - G\mathbf{x}) \leq 0\}$ with the relaxed condition $\|\mathbf{x}\| \leq 1, \|\lambda\| \leq 1$, giving us a convex constraint. To proceed, we will use the following theorem.

Theorem 3. *(Simplified version of Sion's Minimax Theorem) Let U, V be two convex compact subset of $\mathbb{R}^m$ If f is a continuous real-valued function on $U \times V$ with:*

1. *For all $x_1, x_2 \in U$, $y \in V$, $t \in [0,1]$, $f(tx_1 + (1-t)x_2, y) \geq \min\{f(x_1,y), f(x_2,y)\}$*
2. *For all $y_1, y_2 \in V$, $x \in U$, $t \in [0,1]$, $f(x, ty_1 + (1-t)y_2) \leq \max\{f(x,y_1), f(x,y_2)\}$*

Then

$$\min_{x \in U} \max_{y \in V} f(x,y) = \max_{y \in V} \min_{x \in U} f(x,y).$$

Fixing one of λ and $\mathbf{x}$, $\lambda^T(\kappa\mathbf{1} - G\mathbf{x})$ is an affine linear transformation, so it satisfies the condition in the above thoerem. Therefore

$$\min_{\mathbf{x}} \max_{\lambda \geq 0} \lambda^T(\kappa\mathbf{1} - G\mathbf{x}) = \max_{\lambda \geq 0} \min_{\mathbf{x}} \lambda^T(\kappa\mathbf{1} - G\mathbf{x})$$

which implies

$$-\min_{\mathbf{x}} \max_{\lambda \geq 0} \lambda^T(\kappa\mathbf{1} - G\mathbf{x}) = \min_{\lambda \geq 0} \max_{\mathbf{x}} \lambda^T(G\mathbf{x} - \kappa\mathbf{1}).$$

A is non-empty if and only if $\min_{\lambda \geq 0} \max_{\mathbf{x}} \lambda^T(G\mathbf{x} - \kappa\mathbf{1}) \geq 0$. To compute a probabilistic bound for such event, we start with the following inequality (We assume the constraints $\|\lambda\| \leq 1, \|\mathbf{x}\| \leq 1$). Let $g, \mathbf{g}, \mathbf{h}$ be defined the same way as in the last section, then for any $\varepsilon > 0$, we have:

$$\mathbb{P}\Big(\min_{\lambda\geq 0}\max_{\mathbf{x}} \lambda^T G\mathbf{x} + g\|\mathbf{x}\|\|\lambda\| - \lambda^T \mathbf{1}\kappa - \varepsilon\sqrt{n}\|\mathbf{x}\|\|\lambda\| \geq 0\Big) \leq \mathbb{P}\Big(\min_{\lambda\geq 0}\max_{\mathbf{x}} \lambda^T G\mathbf{x} - \lambda^T \mathbf{1}\kappa \geq 0\Big) + \mathbb{P}(g \geq \varepsilon\sqrt{n}) \quad (8)$$

Again the second term on the left-hand side is bounded by $e^{-\varepsilon^2 n/2}$. We focus on the left-hand side. $\lambda_i G\mathbf{x}_j + g\|\lambda_i\|\|\mathbf{x}_j\|$ will be the Y_{ij} in theorem 2. $\|\mathbf{x}_i\|\mathbf{g}^T\lambda_j + \|\lambda_j\|\mathbf{h}^T\mathbf{x}_i$ will be the X_{ij} in theorem 2.Now let's check the conditions in theorem 2. For each i, j, l, k

$$\mathbb{E}(\lambda_i G\mathbf{x}_j + g\|\lambda_i\|\|\mathbf{x}_j\|)(\lambda_l G\mathbf{x}_k + g\|\lambda_l\|\|\mathbf{x}_k\|) = (\lambda_i^T\lambda_l)(\mathbf{x}_j^T\mathbf{x}_k) + \|\lambda_i\|\|\mathbf{x}_j\|\|\lambda_l\|\|\mathbf{x}_k\|$$

$$\mathbb{E}(\|\mathbf{x}_j\|\mathbf{g}^T\lambda_i + \|\lambda_i\|\mathbf{h}^T\mathbf{x}_j)(\|\mathbf{x}_k\|\mathbf{g}^T\lambda_l + \|\lambda_l\|\mathbf{h}^T\mathbf{x}_k) = \|\mathbf{x}_j\|\|\mathbf{x}_k\|\lambda_i^T\lambda_l + \|\lambda_i\|\|\lambda_l\|\mathbf{x}_j^T\mathbf{x}_k$$

By similar arguments in section 2.1, it is easy to verify that the conditions in theorem 2 are satisfied. Therefore, by Gordon's inequality,

$$\mathbb{P}\Big(\min_{\lambda\geq 0}\max_{\mathbf{x}} \lambda^T G\mathbf{x} + g\|\mathbf{x}\|\|\lambda\| - \lambda^T \mathbf{1}\kappa - \varepsilon\sqrt{n}\|\mathbf{x}\|\|\lambda\| \geq 0\Big) \geq \mathbb{P}\Big(\min_{\lambda\geq 0}\max_{\mathbf{x}} \|\mathbf{x}\|\mathbf{g}^T\lambda + \|\lambda\|\mathbf{h}^T\mathbf{x} - \lambda^T \mathbf{1}\kappa - \varepsilon\sqrt{n}\|\mathbf{x}\|\|\lambda\| \geq 0\Big) \quad (9)$$

Solving the maximization problem (terms does not depending on $\mathbf{x}$ will not show below):

$$\begin{aligned}
&\max_{\mathbf{x}} \|\mathbf{x}\|\mathbf{g}^T\lambda + \|\lambda\|\mathbf{h}^T\mathbf{x} - \varepsilon\sqrt{n}\|\mathbf{x}\|\|\lambda\| \\
&= \max_{\mathbf{x}} \|\mathbf{x}\|\Big(\mathbf{g}^T\lambda - \varepsilon\sqrt{n}\|\lambda\| + \frac{\|\lambda\|}{\|\mathbf{x}\|}\mathbf{h}^T\mathbf{x}\Big) \\
&= \max_{\|\mathbf{x}\|} \|\mathbf{x}\|\Big(\mathbf{g}^T\lambda - \varepsilon\sqrt{n}\|\lambda\| + \|\lambda\|\|\mathbf{h}\|\Big) \\
&= \max\{0, \mathbf{g}^T\lambda - \varepsilon\sqrt{n}\|\lambda\| + \|\lambda\|\|\mathbf{h}\|\} \\
&\geq \mathbf{g}^T\lambda - \varepsilon\sqrt{n}\|\lambda\| + \|\lambda\|\|\mathbf{h}\|
\end{aligned}$$

So we have:

$$\mathbb{P}\Big(\min_{\lambda\geq 0}\max_{\mathbf{x}} \|\mathbf{x}\|\mathbf{g}^T\lambda + \|\lambda\|\mathbf{h}^T\mathbf{x} - \lambda^T \mathbf{1}\kappa - \varepsilon\sqrt{n}\|\mathbf{x}\|\|\lambda\| \geq 0\Big) \geq \mathbb{P}\Big(\min_{\lambda\geq 0} \mathbf{g}^T\lambda - \lambda^T \mathbf{1}\kappa - \varepsilon\sqrt{n}\|\lambda\| + \|\lambda\|\|\mathbf{h}\| \geq 0\Big) \quad (10)$$

Now we minimize $\mathbf{g}^T\lambda - \lambda^T\mathbf{1}\kappa - \varepsilon\sqrt{n}\|\lambda\| + \|\lambda\|\|\mathbf{h}\|$ over the feasible region of λ:

$$\begin{aligned}
&\min_{\lambda\geq 0}\mathbf{g}^T\lambda - \lambda^T\mathbf{1}\kappa - \varepsilon\sqrt{n}\|\lambda\| + \|\lambda\|\|\mathbf{h}\| \\
&= \min_{\lambda\geq 0}\|\lambda\|\Big(\frac{\lambda^T}{\|\lambda\|}(\mathbf{g}-\mathbf{1}\kappa) - \varepsilon\sqrt{n} + \|\mathbf{h}\|\Big) \\
&= \min_{\lambda\geq 0}\|\lambda\|\Big(-\|(\mathbf{1}\kappa-\mathbf{g})_+\| - \varepsilon\sqrt{n} + \|\mathbf{h}\|\Big)
\end{aligned}$$

Therefore:

$$\begin{aligned}
&\mathbb{P}\Big(\min_{\lambda\geq 0}\mathbf{g}^T\lambda - \lambda^T\mathbf{1}\kappa - \varepsilon\sqrt{n}\|\lambda\| + \|\lambda\|\|\mathbf{h}\| \geq 0\Big) \\
&= \mathbb{P}\Big[\min_{\lambda\geq 0}\|\lambda\|\Big(-\|(\mathbf{1}\kappa-\mathbf{g})_+\| - \varepsilon\sqrt{n} + \|\mathbf{h}\|\Big) \geq 0\Big] \\
&= \mathbb{P}(-\|(\mathbf{1}\kappa-\mathbf{g})_+\| - \varepsilon\sqrt{n} + \|\mathbf{h}\| \geq 0) \\
&= \mathbb{P}\left(-\sqrt{\sum_{i=1}^m(\kappa-g_i)_+^2} + \sqrt{\sum_{i=1}^n h_i^2} - \varepsilon\sqrt{n} \geq 0\right)
\end{aligned}$$

Notice that $(\kappa-g)_+^2$ and $(\kappa+g)_+^2$ have the same distribution. By our assumption, we can find $\varepsilon>0$ such that

$$\sqrt{\alpha\mathbb{E}(\kappa-g)_+^2} - 1 + 3\varepsilon < 0 \implies -\sqrt{m\mathbb{E}(\kappa-g)_+^2} + \sqrt{n} - 2\varepsilon\sqrt{n} > \varepsilon\sqrt{n}.$$

This implies

$$\begin{aligned}
&\mathbb{P}\left(-\sqrt{\sum_{i=1}^m(\kappa-g_i)_+^2} + \sqrt{\sum_{i=1}^n h_i^2} - \varepsilon\sqrt{n} \geq 0\right) \\
&\geq \mathbb{P}\left(-\sqrt{\sum_{i=1}^m(\kappa-g_i)_+^2} + \sqrt{\sum_{i=1}^n h_i^2} \geq -\sqrt{m\mathbb{E}(\kappa-g)_+^2} + \sqrt{n} - 2\varepsilon\sqrt{n}\right) \\
&= 1 - \mathbb{P}\left(-\sqrt{\sum_{i=1}^m(\kappa-g_i)_+^2} + \sqrt{\sum_{i=1}^n h_i^2} < -\sqrt{m\mathbb{E}(\kappa-g)_+^2} + \sqrt{n} - 2\varepsilon\sqrt{n}\right) \\
&\geq 1 - \mathbb{P}\left(\sqrt{\sum_{i=1}^m(\kappa-g_i)_+^2} > \sqrt{m\mathbb{E}(\kappa-g)_+^2} + \varepsilon\sqrt{n}\right) - \mathbb{P}\left(\sqrt{\sum_{i=1}^n h_i^2} \leq \sqrt{n} - \varepsilon\sqrt{n}\right)
\end{aligned} \tag{11}$$

By similar arguments in section 2.1, there exists $c>0$, such that

$$\mathbb{P}\Big(\min_{\lambda\geq 0}\max_{\mathbf{x}}\lambda^T G\mathbf{x} - \lambda^T\mathbf{1}\kappa \geq 0\Big) \geq 1 - e^{-cn}.$$

Therefore,

$$\mathbb{P}(A \text{ is not empty}) \geq 1 - e^{-cn}$$

References

1. Stojnic, M.: Another Look at the Gardner Problem, 2013
2. Professor Panchenko's lecture notes