



HAL
open science

Main result on self concordant Perceptron.

Adrien Chan-Hon-Tong

► To cite this version:

| Adrien Chan-Hon-Tong. Main result on self concordant Perceptron.. 2020. hal-00722920v34

HAL Id: hal-00722920

<https://hal.science/hal-00722920v34>

Preprint submitted on 3 May 2021 (v34), last revised 16 Jan 2023 (v38)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Main result on self concordant Perceptron.

Adrien CHAN-HON-TONG

May 3, 2021

Abstract

This paper is a summary of paper hal-02491694 focusing on the main theorem.

1 Claim

The paper hal-02491694 focuses on the problem of:

Finding x such that $Ax > \mathbf{0}$ for a given matrix $A \in \mathbb{Z}^{M \times N}$ with the prior that one solution to this set of strict inequality exists.

which is linked with general linear programming (see the complete paper for details).

definition: Let introduce the self concordant function:

$$F(v) = \frac{v^T AA^T v}{2} - \mathbf{1}^T \log(v) = \sum_{i,j=1}^M v_i v_j \times A_i A_j^T - \sum_{m=1}^M \log(v_m) \quad (1)$$

theorem: $A \in \mathbb{Q}^{M \times N}$ a normalized linear feasibility instance with $Ax \geq \mathbf{1}$

- $F(\frac{1}{M}\mathbf{1}) \leq 1 + M \log(M)$
- F has a minimum (let write it F^*) with $-F^* \leq M \log(x^T x)$
- for all v , $F(v) - F^* \leq \frac{1}{2Mx^T x + 2} \Rightarrow AA^T v > \mathbf{0}$

trivial corollary: As F is self concordant [3], damped Newton descent, starting from any v_{start} , builds v such that $F(v) - F^* \leq \frac{1}{2Mx^T x + 2}$ in less than $\tilde{O}(F(v_{start}) - F^* + \log \log(2Mx^T x))$ steps. But, the main theorem gives exactly the required bound: damped Newton descent, starting from $\frac{1}{M}\mathbf{1}$, builds a solution of the linear feasibility in less than $\tilde{O}(M \log(x^T x) + M \log(M) + \log \log(2Mx^T x)) = \tilde{O}(M \log(x^T x))$ steps.

Yet, as for the Perceptron base method [5, 4], $\log(x^T x) \leq L$ where L is the binary size of the input matrix. This finally gives a $\tilde{O}(ML)$ number of steps, with each step being the resolution of a $M \times M$ linear system (which can be done in M^ω with $\omega \approx 2.38$ with [1]). Thus, arithmetic time complexity (operations on \mathbb{Q} are counted as 1) of this algorithm is $\tilde{O}(M^\omega ML)$ which is higher than $\tilde{O}(M^\omega \sqrt{ML})$ for log barrier [3] or even $\tilde{O}((M^\omega + M^{\frac{13}{6}})L)$ for [2] (contrary to what was claimed in version prior than 8 in hal-02491694, L in complexity analysis should not be interpreted as NMB thus arithmetic time complexity of this algorithm is $\tilde{O}(M^\omega ML)$ and not $\tilde{O}(M^\omega L)$). However, binary properties of this algorithm are interesting (see hal-02491694).

2 Proof of the claim

2.1 Existence of a minimum

Let recall Cauchy inequality: $\forall u, v$ two vectors $u^T v \leq \sqrt{u^T u \times v^T v}$.

As A is normalized, then for all m , $A_m A_m^T = 1$. So from Cauchy, $\forall i, j$, $A_i A_j^T \leq 1$. So, $F(\frac{1}{M}\mathbf{1}) = \frac{1}{M^2} \sum_{i,j} A_i A_j^T + M \log(M) \leq 1 + M \log(M)$.

Then, again from Cauchy $(A^T v)^T x \leq \sqrt{v^T A A^T v \times x^T x}$. But $(A^T v)^T x = v^T (Ax)$. And, by definition $Ax \geq \mathbf{1}$. Injecting this last inequality is interesting when $v \geq \mathbf{0}$: $\forall v \geq \mathbf{0}$, $v^T \mathbf{1} \leq v^T (Ax) = (A^T v)^T x \leq \sqrt{v^T A A^T v \times x^T x}$.

Also when $v \geq \mathbf{0}$, then, both $v^T (Ax) = (A^T v)^T x$ and $\sqrt{v^T A A^T v \times x^T x}$ are positive. So, one can take the square: $\forall v \geq \mathbf{0}$, $\frac{(v^T \mathbf{1})^2}{x^T x} \leq v^T A A^T v$. And, independently $(v^T \mathbf{1})^2 > v^T v$ because $v \geq \mathbf{0}$. So $\forall v \geq \mathbf{0}$, $\frac{v^T v}{x^T x} \leq v^T A A^T v$.

Let introduce $f(t) = \frac{t^2}{2x^T x} - \log(t)$, from previous inequality it stands that $F(v) \geq \sum_m f(v_m)$.

Now, f is a single variable function which goes to infinity when t goes to 0 ($t^2 \rightarrow 0$ but $-\log(t) \rightarrow \infty$) or to infinity (t^2 grows faster than $\log(t)$). So, f has a minimum and so F too. Let call them f^* and F^* .

As f and F are smooth, the minimums are characterized by a null derivative or gradient. $f'(t) = \frac{t}{x^T x} - \frac{1}{t}$, so, $f'(\sqrt{x^T x}) = 0$, so $f^* = f(\sqrt{x^T x}) = \frac{1}{2} - \frac{1}{2} \log(x^T x) \geq -\log(x^T x)$. Thus, the minimum of F verifies $F^* \geq M f^* \geq -M \log(x^T x)$.

So the two first assertions of the main theorem are proven.

2.2 Normalization, linearization and lemmas

Independently, let remark that $\theta(t) = F(tv) = \frac{v^T A A^T v}{2} t^2 - \mathbf{1}^T \log(v) - M \log(t)$ is minimal when $v^T A A^T v = M$. So for any w , one could build a $v = \mu w$ such that $v^T A A^T v = M$ and $F(v) \leq F(w)$. In other words, it stands that $F\left(\sqrt{\frac{M}{v^T A A^T v}} v\right) \leq F(v)$.

So, let consider $v \geq \mathbf{0}$ such that $v^T A A^T v = M$. As, $v^T A A^T v \geq \frac{(\mathbf{1}v)^2}{x^T x}$, no v_m could be higher than $\sqrt{M x^T x}$ i.e. $\mathbf{0} \leq v \leq \sqrt{M x^T x} \mathbf{1}$.

Let also remark that $F(v+w) = \frac{v^T A A^T v}{2} + \frac{w^T A A^T w}{2} + w^T A A^T v - \mathbf{1}^T \log(v) - \mathbf{1}^T \log(1 + \frac{w}{v}) = F(v) + \frac{w^T A A^T w}{2} + w^T A A^T v - \mathbf{1}^T \log(1 + \frac{w}{v})$

Finally, let consider the following lemmas from basic analysis:

1. $\phi(t) = \frac{1}{2} \alpha t^2 - \log(1+t) \leq \frac{1}{2} (\alpha+1) t^2 - t = \psi(t)$ for $t \geq 0$
2. $\psi(\frac{1}{\alpha+1}) \leq -\frac{1}{2\alpha+2}$
3. $\phi(\frac{1}{\alpha+1}) \leq -\frac{1}{2\alpha+2}$ i.e. $\forall \alpha \geq 0$, $\frac{1}{2} \frac{\alpha}{(\alpha+1)^2} - \log(1 + \frac{1}{\alpha+1}) \leq -\frac{1}{2\alpha+2}$

Lemma: $\psi'(t) - \phi'(t) = (\alpha+1)t - 1 - \alpha t + \frac{1}{1+t} = t - 1 + \frac{1}{1+t} = \frac{t^2}{1+t} > 0$, so $\psi(t) - \phi(t)$ always increases. But, $\psi(0) = \phi(0) = 0$ so $\psi(t) \geq \phi(t)$ for

$t \geq 0$. Lemma2: $\psi(\frac{1}{\alpha+1}) = \frac{1}{2}(\alpha+1)\frac{1}{(\alpha+1)^2} - \frac{1}{\alpha+1} = -\frac{1}{2\alpha+2}$. lemma3 is just lemma1+lemma2.

2.3 Convergence

Now, either $AA^T v > \mathbf{0}$ (problem solved) or there exists k such that $A_k A_k^T v \leq 0$.

Let consider this case $A_k A_k^T v \leq 0$ and $v^T AA^T v = M$, and, let introduce $w = v + \frac{vk}{v_k^2+1} \mathbf{1}_k$.

Then $F(w) = F(v + \frac{vk}{v_k^2+1} \mathbf{1}_k) = F(v) + \frac{A_k A_k^T}{2} (\frac{vk}{v_k^2+1})^2 + A_k A_k^T v \times \frac{vk}{v_k^2+1} - \log(1 + \frac{1}{v_k^2+1})$. But, $A_k A_k^T v \leq 0$ (by assumption) and $A_k A_k^T = 1$, so $F(w) \leq F(v) + \frac{1}{2} (\frac{vk}{v_k^2+1})^2 - \log(1 + \frac{1}{v_k^2+1})$.

And, from lemmas just above (consider $\alpha = v_k^2$), $F(w) \leq F(v) - \frac{1}{2v_k^2+2}$.

But, $v_k \leq \sqrt{Mx^T x}$, so, $F(w) \leq F(v) - \frac{1}{2Mx^T x+2}$ which is impossible if $F(v) - F^* < \frac{1}{2Mx^T x+2}$. So, $\forall v > \mathbf{0}$ such that $v^T AA^T v = M$, $F(v) - F^* \leq \frac{1}{2Mx^T x+2} \Rightarrow AA^T v > \mathbf{0}$.

Finally, the requirement that $v^T AA^T v = M$ could be remove because normalizing decreases F : $\forall v > \mathbf{0}$, $F(v) - F^* \leq \frac{1}{2Mx^T x+2} \Rightarrow F(\sqrt{\frac{M}{v^T AA^T v}} v) - F^* \leq \frac{1}{2Mx^T x+2} \Rightarrow \sqrt{\frac{M}{v^T AA^T v}} \times AA^T v > \mathbf{0} \Rightarrow AA^T v > \mathbf{0}$.

References

- [1] Andris Ambainis, Yuval Filmus, and François Le Gall. Fast matrix multiplication: limitations of the coppersmith-winograd method. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*, pages 585–593, 2015.
- [2] Michael B Cohen, Yin Tat Lee, and Zhao Song. Solving linear programs in the current matrix multiplication time. In *Proceedings of the 51st annual ACM SIGACT symposium on theory of computing*, 2019.
- [3] Arkadi Nemirovski. Interior point polynomial time methods in convex programming. *Lecture notes*, 2004.
- [4] Javier Peña and Negar Soheili. A deterministic rescaled perceptron algorithm. *Mathematical Programming*, 155(1-2):497–510, 2016.
- [5] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 1958.