Presented as late breaking results contribution
21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A Polynomial Time Algorithm For Minimizing Strongly Convex Functions With Strongly Convex Constraints

Costandin Marius * **Costandin Beniamin** ** **Petru Dobra** ***

* *General Digits(e-mail: costandinmarius@gmail.com)*
** *Technical University of Cluj Napoca(e-mail: bcostandin@yahoo.com)*
*** *Technical University of Cluj Napoca, (e-mail: petru.dobra@aut.utcluj.ro)*

Abstract: We present a novel feasibility criteria for the intersection of finitely many convex sets where each set is given by an inequality. This criteria allows us to easily assert the feasibility by analyzing the unconstrained minimum of a certain convex function, that we form with the given sets. Such a criteria is then used together with bisection techniques to solve general convex optimization problems within a desired precision in polynomial time. A simple complexity analysis is given for the case where the functions involved are strongly convex, but the method can be used for general convex functions.

Keywords: convex optimization, strongly convex function, bisection, feasibility

1. INTRODUCTION

This paper presents an algorithm for answering a very important question often arising in science, engineering and philosophy: the question of feasibility. Given more constraints one may often wonder if it is possible to meet all of them. This approach can be used as a basis for optimization algorithms, therefore we search for a fast way of deciding upon feasibility of different constraints. We give a sufficient criteria for asserting the infeasibility based on the results of an unconstrained optimization of a proposed function. In the case of strongly convex functions we show that the answer can be obtained fast.

Recall the following well known operator:

$$\frac{\partial}{\partial X} = \begin{bmatrix} \frac{\partial}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} \end{bmatrix} = \nabla^T \tag{1}$$

1.1 Convex domains of interest

Let $X \in \mathbb{R}^{n\times 1}$, $n, m \in \mathbb{N}_+$ and let $g_k : \mathbb{R}^{n\times 1} \to \mathbb{R}$ be convex functions for $k \in \{1, \dots, m\}$. Then we define the convex sets:

$$S_k = \left\{ X \in \mathbb{R}^{n\times 1} \middle| g_k(X) \leq 0 \right\} \tag{2}$$

and we are interested if the set

$$S = \bigcap_{k=1}^{m} S_k \tag{3}$$

is empty or not.

Let us define the function $G : \mathbb{R}_+ \times \mathbb{R}^{n\times 1} \to \mathbb{R}_+$

$$G(\alpha, X) = \sum_{k=1}^{m} \log\left(1 + e^{\alpha\cdot g_k(X)}\right) \tag{4}$$

and check for the convexity of $G(\alpha, X)$ by evaluating its derivatives:

$$\frac{\partial G}{\partial X} = \sum_{k=1}^{m} \frac{e^{\alpha\cdot g_k(X)}}{1 + e^{\alpha\cdot g_k(X)}} \cdot \alpha \cdot \frac{\partial g_k}{\partial X} \tag{5}$$

and

$$\frac{\partial}{\partial X}^T \frac{\partial}{\partial X} G = \sum_{k=1}^{m} \frac{e^{\alpha\cdot g_k(X)}}{\left(1 + e^{\alpha g_k(X)}\right)^2} \cdot \alpha^2 \cdot \frac{\partial g_k}{\partial X}^T \frac{\partial g_k}{\partial X} +$$
$$+ \sum_{k=1}^{m} \frac{e^{\alpha\cdot g_k(X)}}{1 + e^{\alpha\cdot g_k(X)}} \cdot \alpha \cdot \frac{\partial}{\partial X}^T \frac{\partial g_k}{\partial X} \tag{6}$$

It is easy to see that the hessian of G is positively defined hence G is convex.

Remark 1. An important class of convex sets is generated by the following system of linear inequalities:

$$A \cdot X + B \preceq 0_m \tag{7}$$

with $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times 1}, X \in \mathbb{R}^{n\times 1}$ and $m, n \in \mathbb{N}$. These are m inequalities of the following form: $g_i(X) = A_i^T \cdot X + b_i \leq 0$ where A_i^T is the i'th line of the matrix A and b_i is the i'th element of the vector B.

At the very core of the following work it is the minimization of the above defined function G, for different parameters $\alpha > 0$ to an ϵ precision $G(\alpha, X) - G(\alpha, X^\star) \leq \epsilon$. The complexity analysis of the given algorithms will account for the number of minimizations required to reach a desired conclusion. Therefore in the following subsection we give some known results for the minimization of a convex function.

1.2 Results concerning unconstrained minimization of convex functions

Let us now give a result concerning the minimization of strongly convex functions. The results are well known but we provide a different proof to the classical one, Robert

M. Gower (2018). Our proof can also be used to show the stability of certain dynamical systems.

Let $F : \mathbb{R}^{n\times 1} \to \mathbb{R}$ be a convex function and consider the dynamical system given below:

$$\dot{X}(t)^T = -\frac{\partial F}{\partial X}(X(t)) \tag{8}$$

Then multiplying at the right by $\dot{X}$ one obtains:

$$\|\dot{X}(t)\|^2 = \dot{X}^T\dot{X} = -\frac{\partial F}{\partial X}\cdot\dot{X} = -\frac{d}{dt}F(X(t)) \tag{9}$$

Let us give the following lemma:

Lemma 2. Let $f(t) = F(X(t))$, then f is convex for all $t \in \mathbb{R}_+$.

Proof. Indeed

$$\begin{aligned} f'(t) &= \frac{\partial F}{\partial X}\cdot\dot{X} = -\dot{X}^T\dot{X} \le 0 \\ f''(t) &= \frac{d}{dt}\left(\frac{\partial F}{\partial X}\cdot\dot{X}\right) = \left(\frac{d}{dt}\frac{\partial F}{\partial X}\cdot\dot{X} + \frac{\partial F}{\partial X}\cdot\ddot{X}\right) \\ &= -2\cdot\ddot{X}^T\cdot\dot{X} \end{aligned} \tag{10}$$

where

$$\begin{aligned} -\ddot{X}^T &= \left[\tfrac{d}{dt}\tfrac{\partial F}{\partial x_1}\ \dots\ \tfrac{d}{dt}\tfrac{\partial F}{\partial x_n}\right] \\ &= \left[\left(\tfrac{\partial}{\partial X}\tfrac{\partial F}{\partial x_1}\right)\cdot\dot{X}\ \dots\ \left(\tfrac{\partial}{\partial X}\tfrac{\partial F}{\partial x_n}\right)\cdot\dot{X}\right] \\ &= \left[\dot{X}^T\cdot\left(\tfrac{\partial}{\partial X}\tfrac{\partial F}{\partial x_1}\right)^T\ \dots\ \dot{X}^T\cdot\left(\tfrac{\partial}{\partial X}\tfrac{\partial F}{\partial x_n}\right)^T\right] = \\ &= \dot{X}^T\cdot\left[\left(\tfrac{\partial}{\partial X}\tfrac{\partial F}{\partial x_1}\right)^T\ \dots\ \left(\tfrac{\partial}{\partial X}\tfrac{\partial F}{\partial x_n}\right)^T\right] = \dot{X}^T\cdot H(X(t)) \end{aligned} \tag{11}$$

where $H = \left(\frac{\partial}{\partial X}\right)^T\frac{\partial F}{\partial X}$ is the Hessian matrix of F. In conclusion $f''(t) = 2\cdot\dot{X}(t)^T\cdot H(X(t))\cdot\dot{X}(t) \ge 0$ since the hessian of F is positively defined, therefore $f(t)$ is convex.

Because we assume a numeric integration of the dynamical system (8) with a constant step size, we will consider in the following the time required to reach the equilibrium point, which therefore will be proportional with the number of integration steps, as a measure of algorithm computational complexity.

Under the hypothesis that F is strongly convex, results about the velocity of convergence can be obtained in Lemma 3:

Lemma 3. (Minimization of strongly convex functions). For a strongly convex function, $F : \mathbb{R}^n \to \mathbb{R}$, the solution to the dynamical system (8), $X(t)$, meets $\|F(X(t)) - F(X(\infty))\| \le \epsilon$ if

$$t \ge \frac{1}{2\cdot K}\cdot\log\left(\frac{\left\|\frac{\partial F}{\partial X}(X(0))\right\|^2}{\epsilon\cdot 2\cdot K}\right) \tag{12}$$

where $K > 0$ such that $\frac{\partial}{\partial X}^T\frac{\partial}{\partial X}F(X) - K\cdot\mathcal{I} \succeq 0$

Proof. Assuming that $\exists K > 0$ such that $\dot{X}^T\cdot H(X)\cdot\dot{X} - K\cdot\dot{X}^T\cdot\dot{X} \ge 0$ (i.e F is strongly convex) one obtains:

$$f'' + 2\cdot K\cdot f' = \dot{X}^T\cdot(2\cdot H(X) - 2\cdot K\cdot\mathcal{I})\cdot\dot{X} \ge 0 \tag{13}$$

hence using Gronwall's inequality lemma one obtains:

$$f'(t) \ge f'(0)\cdot e^{-2\cdot K\cdot t} \iff 0 \le -f'(t) \le -f'(0)\cdot e^{-2\cdot K\cdot t} \tag{14}$$

which means that for a given $X_0 = X(0)$ one obtains

$$0 \le \dot{X}^T(t)\cdot\dot{X}(t) \le \left\|\frac{\partial F}{\partial X}(X_0)\right\|^2\cdot e^{-2\cdot K\cdot t} \tag{15}$$

From Robert M. Gower (2018) since F is strongly convex

$$\begin{aligned} F(X(t+T)) \ge &F(X(t)) + \frac{\partial F}{\partial X}(X(t))\,(X(t+T) - X(t)) + \\ &+ \frac{K}{2}\|X(t+T) - X(t)\|^2 \end{aligned} \tag{16}$$

therefore, after some algebraic manipulations, one obtains:

$$\begin{aligned} &F(X(t)) - F(X(t+T)) \le \\ &-\left\|\sqrt{\frac{K}{2}}\,(X(t) - X(t+T)) - \frac{1}{\sqrt{2K}}\frac{\partial F}{\partial X}(X(t))\right\|^2 \\ &+ \frac{1}{2\cdot K}\left\|\frac{\partial F}{\partial X}(X(t))\right\|^2 \end{aligned} \tag{17}$$

therefore, letting $T \to \infty$

$$\begin{aligned} \left|F(X(t)) - F(X^\star)\right| &\le \frac{1}{2\cdot K}\cdot\left\|\frac{\partial F}{\partial X}(X(t))\right\|^2 = \frac{-f'(t)}{2\cdot K} \\ &\le \frac{e^{-2\cdot K\cdot t}}{2\cdot K}\cdot\left\|\frac{\partial F}{\partial X}(X_0)\right\|^2 \le \epsilon \end{aligned} \tag{18}$$

from where the conclusion easily follows.

Remark 4. Let us assume that the dynamical system is integrated using Euler technique:

$$X(t_{p+1}) = X(t_p) - T_s\cdot\frac{\partial F}{\partial X}^T(X(t_p)) \tag{19}$$

where T_s is the sampling time and $t_p = p\cdot T_s$. In order to achieve (12) one has

$$p \ge \frac{1}{K\cdot T_s}\cdot\log\left(\frac{\left\|\frac{\partial F}{\partial X}(X(0))\right\|^2}{\epsilon\cdot 2\cdot K}\right) \tag{20}$$

hence minimizing a strongly convex function to ϵ precision takes $\mathcal{O}(\log\left(\frac{\left\|\frac{\partial F}{\partial X}(X(0))\right\|^2}{\epsilon}\right))$ steps constant along gradient descent direction.

2. MAIN RESULTS

2.1 An Algorithm For Asserting The Feasibility of An Intersection of Convex Sets

Let us consider the following system of convex inequalities:

$$g_k(X) \le 0 \tag{21}$$

with $k \in \{1,\dots,m\}, X \in \mathbb{R}^{n\times 1}$ and $m, n \in \mathbb{N}$. We want to solve the feasibility problem related to the above system of inequalities.

Definition 5. We say that the system of inequalities 21 is $\delta \in \mathbb{R}$ feasible, if $\exists X \in \mathbb{R}^{n\times 1}$ such that $\forall k \in \{1,\dots,m\}$ one has $g_k(X) \le \delta$.

Let $\epsilon, \delta > 0$ be arbitrarily small. **In the following we shall derive an algorithm which: finds a ϵ feasible point or returns a proof that the system is not $-\delta$ feasible.**

Algorithm derivation Let us recall the function, $G : \mathbb{R}_+ \times \mathbb{R}^{n\times 1} \to \mathbb{R}_+$ from 4

$$G(\alpha, X) = \sum_{k=1}^{m} \log\left(1 + e^{\alpha\cdot g_k(X)}\right) \tag{22}$$

where g_k is a convex function.

Let $(\alpha_p)_{p\in\mathbb{N}} \subset \mathbb{R}_+$ be an increasing, unbounded sequence of positive real numbers and let $(G_p)_{p\in\mathbb{N}} \subset \mathbb{R}$ be a sequence of numbers defined as follows:

$$G_p = \inf\left\{G(\alpha_p, X) | X \in \mathbb{R}^{n\times 1}\right\} \tag{23}$$

Here we give an important lemma, which is the base of an infeasibility criterion:

Lemma 6. If the convex system 21 is $-\delta$ with $\delta > 0$ feasible, then $G_{p+1} \le \eta \cdot G_p$ for some $\eta \in (0, \frac{1}{3})$ and $p \in \mathbb{N}$ if

$$\alpha_{p+1} \ge \frac{-1}{\delta} \cdot \log\left(e^{\frac{\eta\cdot G_p}{m}} - 1\right) \tag{24}$$

Proof. Indeed, if (21) is $-\delta$ feasible, then $\exists \hat{X} \in \mathbb{R}^{n\times 1}$ such that $g_k(X) \le -\delta$ for all k, hence

$$\log\left(1 + e^{\alpha_{p+1}\cdot g_k(X)}\right) \le \log\left(1 + e^{-\alpha_{p+1}\cdot\delta}\right) \tag{25}$$

hence

$$\begin{aligned} G_{p+1} \le \left(m \cdot \log\left(1 + e^{-\alpha_{p+1}\cdot\delta}\right)\right) \le \eta \cdot G_p \iff \\ \alpha_{p+1} \ge \frac{-1}{\delta} \cdot \log\left(e^{\frac{\eta\cdot G_p}{m}} - 1\right) \end{aligned} \tag{26}$$

Remark 7. For each step, while minimizing $G(\alpha_{p+1}, X)$, a *minimizer* may not exist, but we can stop searching for it when

$$|G(\alpha_{p+1}, X) - G_{p+1}| \le \eta \cdot G_p \tag{27}$$

Let X_{p+1} be such a point, hence

$$\begin{aligned} &\left|G(\alpha_{p+1}, X_{p+1}) - \frac{\eta}{2} G_p\right| \le \\ &\le \left|G(\alpha_{p+1}, X_{p+1}) - G_{p+1} + G_{p+1} - \frac{\eta}{2} \cdot G_p\right| \\ &\le |G(\alpha_{p+1}, X_{p+1}) - G_{p+1}| + \left|G_{p+1} - \frac{\eta}{2} \cdot G_p\right| \\ &\le \frac{3\eta}{2} \cdot G_p \end{aligned} \tag{28}$$

Let us denote $\hat{G}_{p+1} = G(\alpha_{p+1}, X_{p+1})$ with X_{p+1} the point found above. Then in the above 6, if G_p is replaced by $\hat{G}_p = G(\alpha_p, X_p)$ in the lower bound for α_{p+1} it can also be replaced in the conclusion, to obtain $G_{p+1} \le \eta \cdot \hat{G}_p$. This together with 28 leads to

$$G(\alpha_{p+1}, X_{p+1}) \le 2 \cdot \eta \cdot G(\alpha_n, X_p) \le \ldots \le (2\cdot\eta)^p \cdot G(\alpha_1, X_1) \tag{29}$$

Remark 8. It is important to remember the implications:

$$\begin{aligned} &\begin{cases} \text{the system is } -\delta \text{ feasible} \\ \alpha_{p+1} \ge \frac{-1}{\delta} \cdot \log\left(e^{\frac{\eta\cdot G_p}{m}} - 1\right) \\ |G(\alpha_{p+1}, X_{p+1}) - G_{p+1}| \le \eta \cdot G(\alpha_p, X_p) \end{cases} \Rightarrow \\ &\Rightarrow G(\alpha_{p+1}, X_{p+1}) \le 2 \cdot \eta \cdot G(\alpha_p, X_p) \end{aligned} \tag{30}$$

If at some step, the right hand side conclusion is not met, it meas that at least one condition from the cases was not true. We will make sure that the last two conditions are met, hence failing to obtain the RHS inequality will mean that the system is not $-\delta$ feasible, for a fixed, initially given $\delta > 0$. This will be a certificate that the system (21) is not $-\delta$ feasible and we **STOP** the feasibility algorithm.

Let $E > 0$ be a given tolerance and let $(2\cdot\eta)^p \cdot G(\alpha_1, X_1) \le E$, then for all k one has

$$\log\left(1 + e^{\alpha_{p+1}\cdot g_k(X)}\right) \le G\left(\alpha_{p+1}, X_{p+1}\right) \le E \tag{31}$$

hence

$$g_k(X) \le \frac{1}{\alpha_{p+1}} \cdot \log\left(e^{\sqrt{E}} - 1\right) = \epsilon \tag{32}$$

where $\epsilon > 0$ is the desired precision, which can be arbitrarily small.

Complexity analysis for strongly convex functions We carry this analysis under the hypothesis that evaluating $g_k(X)$ on a constant requires $\mathcal{O}(n)$ flops, and eventually upon some normalization, the value of g_k on a constant belongs to $\mathcal{O}(1)$ set. Then we consider that evaluating the gradient of $G(\alpha, X)$ requires $\mathcal{O}(m\cdot n)$ flops.

Next, please note that for a given ϵ, η a finite deterministic p, the number of steps, is required to obtain $(2\cdot\eta)^p \cdot G(\alpha_1, X_1) \le E$, therefore $p \in \mathcal{O}(\log(m))$ since $G(\alpha_1, X_1) \in \mathcal{O}(m)$ if $g_k(X_1) \in \mathcal{O}(1)$. Next, each step requires finding X_{p+1} given X_p and δ.

We will carry the analysis for the number of steps required to minimize $G(\alpha, X)$ for the case where g_k are strongly convex functions.

For such functions it is easy to see from (6) that if $g_k(X)$ is strongly convex then let $K > 0$ such that $\frac{\partial}{\partial X}^T \frac{\partial}{\partial X} g_k - K \cdot \mathcal{I} \succeq 0$, hence for all X for which $\exists k$ such that $g_k(X) \ge 0$ one has:

$$\begin{aligned} \frac{\partial}{\partial X}^T \frac{\partial}{\partial X} G_k &\succeq \sum_{k=1}^{m} \frac{e^{\alpha\cdot g_k(X)}}{1 + e^{\alpha\cdot g_k(X)}} \cdot \alpha \cdot \frac{\partial}{\partial X}^T \frac{\partial g_k}{\partial X} \\ &\succeq \frac{\alpha \cdot K}{2} \cdot \mathcal{I} \end{aligned} \tag{33}$$

In terms of the memory required, the parameter $\alpha_p \approx \frac{1}{\delta} \cdot \log(m)$ for very large m.

Since $\left\|\frac{\partial G(\alpha_p, X)}{\partial X}(X_0)\right\| \le m \cdot \alpha_p \cdot \overline{d}_g$ where $\overline{d}_g = \max_k\{\|\frac{\partial g_k}{\partial X}(X_0)\|\} \in \mathcal{O}(1)$, after some eventual normalization, there are $2 \cdot \log(\frac{m\cdot\log(m)}{\epsilon\cdot\delta})$ gradient descent steps required to minimize $G(\alpha_p, X)$ with each step requiring the evaluation of the gradient. Computing the gradient of $G(\alpha_p, X)$ consists of evaluating the $g_k(X)$ for all $k \in \{1, \ldots, m\}$. Under the hypothesis that evaluation of g_k takes n flops one can assume that each gradient descent step requires $\mathcal{O}(m\cdot n)$ flops, hence the overall algorithm requires N flops of where

$$N \in \mathcal{O}\left(m \cdot n \cdot \log\left(\frac{m^2 \cdot \log(m)^2}{\epsilon \cdot \delta^2}\right) \cdot \log(m)\right) \tag{34}$$

for finding a point $X^\star$ such that $g_k(X^\star) \le \epsilon > 0$ or proving that there is no point $X^\star$ such that $g_k(X^\star) \le -\delta < 0$.

2.2 Minimizing convex functions with convex constraints

Having solved the feasibility problem for an intersection of convex sets, it is well known that one can easily proceed to solve a convex minimization problem with convex constraints:

Presented as late breaking results contribution
21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

$$\begin{aligned} \min \quad & f(X) \\ \text{s.t} \quad & g_1(X) \leq 0 \\ & \vdots \\ & g_m(X) \leq 0 \end{aligned} \tag{35}$$

For the sake of the argument we will assume that $f(X) \geq 0$. Assuming the constraints are feasible (otherwise the optimization stops), we find a feasible point, call it X_1 and let $f_1 = f(X_1)$. Next we rewrite (35) like in the following, using the epigraph technique. Let us associate for each convex function g_k forming the restrictions, the function $\hat{g}_k : \mathbb{R}^{n+1} \to \mathbb{R}$

$$\hat{g}_k\left(\hat{X}\right) = g(X) \tag{36}$$

where $\hat{X} = \begin{bmatrix} X \\ x_{n+1} \end{bmatrix}$ for all $X \in \mathbb{R}^n$ in the domain of g_k. Then similarly, let $\hat{f} : \mathbb{R}^{n+1} \to \mathbb{R}$

$$\hat{f}\left(\hat{X}\right) = f(X) - x_{n+1} \tag{37}$$

Finally for $R \in [0, f_1]$ we obtain a family of feasibility problems:

$$\begin{cases} x_{n+1} & \leq R \\ \hat{f}(\hat{X}) & \leq 0 \\ \hat{g}_1(\hat{X}) & \leq 0 \\ & \vdots \\ \hat{g}_m(\hat{X}) & \leq 0 \end{cases} \tag{38}$$

One can start with $R = f_1$ then, if feasible, let $R = \frac{f_1}{2}$ and so on, practically bisecting on the interval $[0, f_1]$. If $f_1 \in \mathcal{O}(1)$, it is well know that in this case, there will be $\log_2\left(\frac{f_1}{\epsilon}\right) \in \mathcal{O}\left(\log\left(\frac{1}{\epsilon}\right)\right)$ steps required to obtain a solution to ϵ precision where the complexity of each step is given by (34).

3. CONCLUSION AND FUTURE WORK

We presented a sufficient criteria for asserting the feasibility of a intersection of convex sets, which proved easy to apply in the case of strongly convex functions. The we merged this into an optimization algorithm based on bisection and epigraph of the function to be minimized. As future work one can try to obtain bounds on the complexity for other convex functions not necessarily strongly convex.

REFERENCES

Stephen Boyd, Lieven Vandenberghe CONVEX ANALYSIS AND NONLINEAR OPTIMIZATION Theory and Examples Springer-Verlag New York 2006

Robert M. Gower (2018) Convergence Theorems for Gradient Descent

Sanjeev Arora, Siddhartha Brahma *The Ellipsoid Algorithm for Linear Programming* COS 521, Fall 2005 Princeton University