Systems Engineering and Computer Science Graduate Program

Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineering

FEDERAL UNIVERSITY OF RIO DE JANEIRO

TECHNICAL REPORT

TR-PESC/Coppe/UFRJ- 798/2025

An Approach to the Pseudo-Huber Function in a Primal-Dual Algorithm

Authors:

Lennin Mallma Ramirez[1], Tran Ngoc Nguyen[2], Nelson Maculan[3]

PESC/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil[1]

DMS, Quy Nhon University, Quy Nhon, Vietnam[2]

PESC/IM/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro Brazil[3]

Corresponding author(s). E-mail(s): lenninmr@cos.ufrj.br

Abstract

In this note, we are interested in solving the minimization problem with equality constraints, with non-convexity assumptions. To solve this problem, we consider the primal-dual algorithm that was studied by Armand and Omheni. But in our approach, we consider the Pseudo-Huber function for the case of equality constraints, and not the quadratic penalty function.

Keywords: Non-convex optimization, Augmented Lagrangian methods, Equality-constrained minimization

Publication Date: 20/08/2025

An Approach to the Pseudo-Huber Function in a Primal-Dual Algorithm

Lennin Mallma Ramirez[1], Tran Ngoc Nguyen[2], Nelson Maculan[3]
PESC/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil[1]
DMS, Quy Nhon University, Quy Nhon, Vietnam[2]
PESC/IM/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro Brazil[3]

Monday 11th August, 2025

Abstract

In this note, we are interested in solving the minimization problem with equality constraints, with non-convexity assumptions. To solve this problem, we consider the primal-dual algorithm that was studied by Armand and Omheni. But in our approach, we consider the Pseudo-Huber function for the case of equality constraints, and not the quadratic penalty function.

1 Introduction and Basic Results

We consider the nonconvex nonlinear programming problem (NLP) as follows

$$(P) \quad \min_{x\in\mathbb{R}^n} f(x), \ \textit{subject to} \quad c(x) = 0, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable functions and not necessarily convex. Inspired by the Primal-Dual algorithms studied in work [1] and work [2], we propose an approach similar to that studied in the mentioned works, but considering a Huber function. We denote by $g(x) \in \mathbb{R}$ the gradient of f at x and by $A(x) \in \mathbb{R}^{n\times m}$ the transpose of the Jacobian matrix of c at x. Let $y \in \mathbb{R}^m$ and $w = (x, y)$. By define

$$F(w) = \begin{pmatrix} g(x) + A(x)y \\ c(x) \end{pmatrix}, \tag{1.2}$$

the first-order optimality conditions of problem (1) are $F(w) = 0$. For an iterate $w_k = (x_k, y_k) \in \mathbb{R}^{n+m}$, $k \in \mathbb{N}$.

1.1 The Pseudo-Huber Primal-Dual Algorithm (PHPDA)

We define the Pseudo-Huber Augmented Lagrangian Function (PHALF) as, $\mathcal{L}_H : \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}_{++} \to \mathbb{R}$, as

$$\begin{aligned}\mathcal{L}(x,\lambda,\sigma) &= f(x) + \sum_{i=1}^{m} \lambda_i c_i(x) + \sum_{i=1}^{m} \frac{1}{\sigma}\left(\sqrt{c_i(x)^2+1} - 1\right) \\ &= f(x) + \sum_{i=1}^{m} \lambda_i c_i(x) + \sum_{i=1}^{m} \frac{1}{\sigma} h\left(c_i(x)\right),\end{aligned} \tag{1.3}$$

where $h : \mathbb{R} \to \mathbb{R}$ is defined as

$$h(t) := \sqrt{t^2+1} - 1.$$

For more details on the Pseudo-Huber function, see work [3]. This function has the following properties:

(a) $h(0) = 0$,

(b) $h'(0) = 0$,

(c) $h'(t) = \frac{t}{\sqrt{t^2+1}}$, with $-1 < h'(t) < 1$, $\forall t \in \mathbb{R}$,

(d) $h''(t) = \frac{1}{\sqrt{(t^2+1)^3}}$, with $0 < h''(t) \leq 1$, $\forall t \in \mathbb{R}$.

We can notice that the h function has properties similar to the hyperbolic function (work [4] and [5]). In the function $\mathcal{L}$, the penalty parameter is $\sigma > 0$ and $\lambda \in \mathbb{R}^m$ is the Lagrange multiplier. The first-order optimality conditions for minimization $\mathcal{L}_H(\cdot,\lambda,\sigma)$ are

$$\nabla\mathcal{L}(x,\lambda,\sigma) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x) + \sum_{i=1}^{m} \frac{1}{\sigma}\left(\frac{c_i(x)}{\sqrt{c_i(x)^2+1}}\right)\nabla c_i(x) = 0, \tag{1.4}$$

factoring properly, we have that

$$\nabla\mathcal{L}(x,\lambda,\sigma) = \nabla f(x) + \sum_{i=1}^{m}\left(\lambda_i + \frac{1}{\sigma}\frac{c_i(x)}{\sqrt{c_i(x)^2+1}}\right)\nabla c_i(x) = 0, \tag{1.5}$$

or, equivalently

$$\nabla\mathcal{L}(x,\lambda,\sigma) = \nabla f(x) + \nabla c(x)^\top\left(\lambda + \frac{1}{\sigma}h'(c(x))\right) = 0, \tag{1.6}$$

where

$$h'(c(x)) = \begin{pmatrix} h'(c_1(x)) & h'(c_2(x)) & \ldots & h'(c_m(x))\end{pmatrix}^\top \in \mathbb{R}^m, \text{ for } x \in \mathbb{R}^n. \tag{1.7}$$

We introducing the dual variable $y \in \mathbb{R}^m$ defined by

$$y_i = \lambda_i + \frac{1}{\sigma}\frac{c_i(x)}{\sqrt{c_i(x)^2+1}}, \quad i = 1,\ldots,m.$$

Let us note the following:

(A1) By (c), we have $-1 < \frac{c_i(x)}{\sqrt{c_i(x)^2+1}} < 1$, then $\sigma\,|y_i - \lambda_i| < 1, \;\; i = 1, ..., m.$

(A2) By (d), we have $0 < \frac{1}{\sigma}\frac{1}{\sqrt{(c_i(x)^2+1)^3}} \leq \frac{1}{\sigma}$.

The optimality conditions can be reformulated as

$$\Phi(w, \lambda, \sigma) := \begin{pmatrix} g(x) + A(x)y \\ F(x)c(x) + \sigma(\lambda - y) \end{pmatrix} = 0, \tag{1.8}$$

where $F(x) = \operatorname{diag}\left(\frac{1}{\sqrt{c_1(x)^2+1}}, \frac{1}{\sqrt{c_2(x)^2+1}}, \ldots, \frac{1}{\sqrt{c_m(x)^2+1}}\right) \in \mathbb{R}^{m \times m}$ is the diagonal matrix, i.e.,

$$F(x) = \begin{pmatrix} \frac{1}{\sqrt{c_1(x)^2+1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{c_2(x)^2+1}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{c_m(x)^2+1}} \end{pmatrix}.$$

The function Φ in the optimality conditions (1.8) can be rewritten under the following form

$$\Phi(w, \lambda, \rho, \sigma) := \begin{pmatrix} g(x) + A(x)y \\ h'(c(x)) + \sigma(\lambda - y) \end{pmatrix}, \tag{1.9}$$

where $h'(c(x))$ is defined in (1.7).

In this case, the regularized Jacobian matrix J of the function Φ with respect to w is defined by

$$J_{\sigma,\theta}(w) = \begin{pmatrix} H_\theta(w) & A(x) \\ D(x)A(x)^\top & -\sigma I \end{pmatrix}, \tag{1.10}$$

where

$$D(x) = \begin{pmatrix} \frac{1}{\sqrt{(c_1(x)^2+1)^3}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{(c_2(x)^2+1)^3}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{(c_m(x)^2+1)^3}} \end{pmatrix} \tag{1.11}$$

$$= \begin{pmatrix} h''(c_1(x)) & 0 & \cdots & 0 \\ 0 & h''(c_2(x)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h''(c_m(x)) \end{pmatrix} \tag{1.12}$$

$$= \operatorname{diag}\left(h''(c_1(x))\; h''(c_2(x)) \ldots h''(c_m(x))\right). \tag{1.13}$$

It is worth noting that the matrix $D(x)$ is positive definite for all $x \in \mathbb{R}^n$.

The second order derivative of the function $\mathcal{L}_H$ with respect to x is

$$\nabla^2 \mathcal{L}_H(x,\lambda,\sigma) = \nabla^2 f(x) + \sum_{i=1}^{m} y_i \nabla^2 c_i(x) + \sum_{i=1}^{m} \frac{1}{\sigma} \frac{1}{\sqrt{(c_i(x)^2+1)^3}} \nabla c_i(x) \nabla c_i(x)^T. \tag{1.14}$$

The primal-dual merit function used in our work is given as below

$$\varphi_{\lambda,\sigma,\nu}(w) = \mathcal{L}_H(x,\lambda,\sigma) + \frac{\nu}{2\sigma} \left\| h'(c(x)) - \sigma(\lambda - y) \right\|^2, \tag{1.15}$$

where $\lambda \in \mathbb{R}^n$ is an estimate of the Lagrange multiplier, $\sigma > 0$ is the penalty parameter and $\nu > 0$ is a scale parameter to balance between two terms in the merit function.

The main of outer and inner iterations is to apply a Newton-type method to solve the nonlinear system (1.8). Hence, at each iteration of these algorithms, algorithms tend to solve the following linear system

$$J_{\sigma,\theta}(w)d = -\Phi(w,\lambda,\sigma), \tag{1.16}$$

where $J_{\sigma,\theta}$ and Φ are respectively defined in (1.10), (1.9) and $d = \begin{pmatrix} d_x & d_y \end{pmatrix} \in \mathbb{R}^{n+n}$ is the direction. This system is equivalent to

$$H(\theta)d_x + A(x)d_y = -\left(g(x) + A(x)y\right) \tag{1.17}$$

$$D(x)A(x)^\top d_x - \sigma d_y = -\left(h'(c(x)) + \sigma(\lambda - y)\right), \tag{1.18}$$

which implies

$$g(x) = -H(\theta)d_x - A(x)(y + d_y) \tag{1.19}$$

$$y + d_y - \frac{1}{\sigma} h'(c(x)) - \lambda = \frac{1}{\sigma} D(x)A(x)^\top d_x. \tag{1.20}$$

From this fact, we have the following results about the descent direction of the merit function (1.15).

Lemma 1.1. *Let $w \in \mathbb{R}^{n+m}$, $\lambda \in \mathbb{R}^n$, $\sigma > 0$, and $\nu > 0$. If $d = (d_x \; d_y)$ is the solution of the linear system (1.16), then*

$$\nabla \varphi_{\lambda,\sigma,\nu}(w)^\top d = -d_x \left(H(w) + \frac{1}{\sigma} A(x)D(x)A(x)^\top \right) d_x - \frac{\nu}{\sigma} \left\| D(x)A(x)^\top d_x - \sigma d_y \right\|^2. \tag{1.21}$$

If $H(w) + (1/\sigma)A(x)D(x)A(x)^\top$ is positive definite and $\Phi(w,\lambda,\sigma)$ is nonzero, then d is a descent direction of the merit function $\varphi_{\lambda,\sigma,\nu}$ at w.

Proof. By taking the gradient of $\varphi_{\lambda,\sigma,\nu}$, we then have

$$\nabla \varphi_{\lambda,\sigma,\nu}(w) = \begin{pmatrix} g(x) + A(x)\lambda + \frac{1}{\sigma} A(x)h'(c(x)) + \frac{\nu}{\sigma} A(x)D(x)(h'(c(x)) + \sigma(\lambda - y)) \\ -\nu(h'(c(x)) + \sigma(\lambda - y)). \end{pmatrix} \tag{1.22}$$

This implies that

$$\begin{aligned}
\nabla \varphi_{\lambda,\sigma,\nu}(w)^\top d &= g(x)^\top d_x + \lambda^\top A(x)^\top d_x + \frac{1}{\sigma} h'(c(x))^\top A(x)^\top d_x \\
&\quad + \frac{\nu}{\sigma}\left(h'(c(x)) + \sigma(\lambda - y)\right)^\top D(x)^\top A(x)^\top d_x - \nu (h'(c(x)) + \sigma(\lambda - y))^\top d_y \\
&= g(x)^\top d_x + \frac{1}{\sigma}\left(h'(c(x)) + \sigma\lambda^\top\right) A(x)^\top d_x \\
&\quad + \frac{\nu}{\sigma}(h'(c(x)) + \sigma(\lambda - y))^\top \left(D(x)^\top A(x)^\top d_x - \sigma d_y\right).
\end{aligned}$$

By substituting (1.19), (1.18) to the above formula with noting that $D(x) = D(x)^\top$ and $H(w) = H(w)^\top$ and then using (1.20), we get

$$\begin{aligned}
\nabla \varphi_{\lambda,\sigma,\nu}(w)^\top d &= \left(-H(w) d_x - A(x)\left(y + d_y\right)\right)^\top d_x + \frac{1}{\sigma}\left(h'(c(x)) + \sigma\lambda^\top\right) A(x)^\top d_x \\
&\quad + \frac{\nu}{\sigma}(D(x)A(x)^\top d_x - \sigma d_y)^\top \left(D(x)A(x)^\top d_x - \sigma d_y\right) \\
&= -d_x^\top H(w) d_x - \left(y + d_y - \frac{1}{\sigma} h'(c(x)) - \lambda^\top\right)^\top A(x)^\top d_x \\
&\quad - \frac{\nu}{\sigma}(D(x)A(x)^\top d_x - \sigma d_y)^\top \left(D(x)A(x)^\top d_x - \sigma d_y\right) \\
&= -d_x^\top H(w) d_x - \left(\frac{1}{\sigma} D(x) A(x)^\top d_x\right)^\top A(x)^\top d_x \\
&\quad - \frac{\nu}{\sigma}\left\|D(x)A(x)^\top d_x - \sigma d_y\right\|^2 \\
&= -d_x^\top \left(H(w) + \frac{1}{\sigma} A(x) D(x) A^\top(x)\right) d_x - \frac{\nu}{\sigma}\left\|D(x)A(x)^\top d_x - \sigma d_y\right\|^2.
\end{aligned}$$

The proof is then complete. □

Funding: This work has been partially supported by FAPERJ (grant E-26/205.684/2022), COPPETEC Foundation, and CNPq.

References

[1] P. Armand and R. Omheni. A globally and quadratically convergent primal–dual augmented lagrangian algorithm for equality constrained optimization. *Optimization Methods and Software*, 32(1):1–21, 2017. 1

[2] P. Armand and N. N. Tran. An augmented lagrangian method for equality constrained optimization with rapid infeasibility detection capabilities. *Journal of Optimization Theory and Applications*, 181:197–215, 2019. 1

[3] K. Fountoulakis and J. Gondzio. A second-order method for strongly convex ℓ 1-regularization problems. *Mathematical Programming*, 156(1):189–219, 2016. 2

[4] L. M. Ramirez, N. Maculan, A. E. Xavier, and V. L. Xavier. Dislocation hyperbolic augmented lagrangian algorithm for nonconvex optimization. *RAIRO-Operations Research*, 57(5):2941–2950, 2023. 2

[5] A. E. Xavier. *Penalização hiperbólica: Um novo método para resolução de problemas de otimização*. PhD thesis, M. Sc. Thesis-COPPE-UFRJ, Rio de Janeiro, 1982. 2