# Cauchy and the Gradient Method

Claude Lemaréchal

2010 Mathematics Subject Classification: 65K05, 90C30
Keywords and Phrases: Unconstrained optimization, descent method, least-square method

Any textbook on nonlinear optimization mentions that the gradient method is due to Louis Augustin Cauchy, in his *Compte Rendu à l'Académie des Sciences* of October 18, 1847[1] (needless to say, this reference takes a tiny place amongst his fundamental works on analysis, complex functions, mechanics, etc. Just have a look at `http://mathdoc.emath.fr/cgi-bin/oetoc?id=OE_CAUCHY_1_10`: a paper every week).

Cauchy is motivated by astronomic calculations which, as everybody knows, are normally very voluminous. To compute the orbit of a heavenly body, he wants to solve *not the differential equations, but the* [*algebraic*] *equations representing the motion of this body, taking as unknowns the elements of the orbit themselves. Then there are six such unknowns.*[2]. Indeed, a motivation related with operations research would have been extraordinary. Yet, it is interesting to note that equation-solving has always formed the vast majority of optimization problems, until not too long ago.

To solve a system of equations in those days, *one ordinarily starts by reducing them to a single one by successive eliminations, to eventually solve for good the resulting equation, if possible. But it is important to observe that 1° in many cases, the elimination cannot be performed in any way; 2° the resulting equation is usually very complicated, even though the given equations are rather simple.*[3] Something else is wanted.

Thus consider a function

$$u = f(x, y, z, \ldots)$$

[1] "Méthode générale pour la résolution des systèmes d'équations simultanées"

[2] non plus aux équations diffŕentielles, mais aux équations finies qui représentent le mouvement de cet astre, et en prenant pour inconnues les éléments mêmes de l'orbite. Alors les inconnues sont au nombre de six.

[3] on commence ordinairement par les réduire à une seule, à l'aide d'éliminations successives, sauf à résoudre définitivement, s'il se peut, l'équation résultante. Mais il importe d'observer, 1° que, dans un grand nombre de cas, l'élimination ne peut s'effectuer en aucune manière ; 2° que l'équation résultante est généralement très-compliquée, lors même que les équations données sont assez simples.

Augustin Louis Cauchy, 1789–1857 (Wikimedia, Cauchy Dibner-Collection Smithsonian Inst.)

of several variables, which never becomes negative, and stays continuous. *To find the values of $x, y, z, \ldots$ satisfying the equation*

$$u = 0\,,$$

*it will suffice to let indefinitely decrease the function u, until it vanishes.*[4]

Start from particular values $\mathrm{x}, \mathrm{y}, \mathrm{z}, \ldots$ of the variables $x, y, z$; call u the corresponding value of $u$ and

$$\mathrm{X} = \mathrm{f}'_x,\ \mathrm{Y} = \mathrm{f}'_y,\ Z = \mathrm{f}'_z,\ \ldots$$

the derivatives.[5] Let $\alpha, \beta, \gamma, \ldots$ be small increments given to the particular values $\mathrm{x}, \mathrm{y}, \mathrm{z}, \ldots$; then there holds approximately

$$f(\mathrm{x}+\alpha, \mathrm{y}+\beta, \mathrm{z}+\gamma, \cdots) = \mathrm{u} + \mathrm{X}\alpha + \mathrm{Y}\beta + \mathrm{Z}\gamma + \cdots\,.$$

Taking $\theta > 0$ and

$$\alpha = -\theta\mathrm{X},\ \beta = -\theta\mathrm{Y},\ \gamma = -\theta\mathrm{Z},\ \ldots\,,$$

we obtain approximately

$$f(\mathrm{x}-\theta\mathrm{X}, \mathrm{y}-\theta\mathrm{Y}, \mathrm{z}-\theta\mathrm{Z}, \ldots) = \mathrm{u} - \theta(\mathrm{X}^2 + \mathrm{Y}^2 + \mathrm{Z}^2 + \cdots)\,. \tag{1}$$

[4]Pour trouver les valeurs de $x, y, z, \ldots$, qui vérifieront l'équation $u = 0$, il suffira de faire décroître indéfiniment la fonction $u$, jusqu'à ce qu'elle s'évanouisse.

[5]Already in those times, one carefully distinguishes a function from *a value* of this function. Observe also that Cauchy cares about continuity but not differentiability ...

*It is easy to conclude that the value* $\Theta$ *of* $u$*, given by the formula*

$$\Theta = f(\mathrm{x} - \theta\mathrm{X}, \mathrm{y} - \theta\mathrm{Y}, \mathrm{z} - \theta\mathrm{Z}, \ldots) \tag{2}$$

*will become smaller than* u *if* $\theta$ *is small enough. If, now,* $\theta$ *increases and if, as we assumed, the function* $f(x, y, z, \cdots)$ *is continuous, the value* $\Theta$ *of* $u$ *will decrease until it vanishes, or at least until it coincides with a* minimal *value, given by the univariate equation*[6]

$$\Theta'_\theta = 0\,. \tag{3}$$

One iteration of the gradient method is thus stated, with two variants: (2) (Armijo-type line-search) or (3) (steepest descent). A third variant, valid when u is already small, is defined by equating (1) to 0:

$$\theta = \frac{\mathrm{u}}{\mathrm{X}^2 + \mathrm{Y}^2 + \mathrm{Z}^2 + \cdots}\,.$$

Other remark: when good approximate values are already obtained, one may switch to Newton's method. Finally, for a system of simultaneous equations

$$u = 0,\ v = 0,\ w = 0,\ \ldots\,,$$

just apply the same idea to the single equation[7]

$$u^2 + v^2 + w^2 + \cdots = 0\,. \tag{4}$$

Convergence is just sloppily mentioned: *If the new value of* u *is not a minimum, one can deduce, again proceeding in the same way, a third value still smaller; and, so continuing, smaller and smaller values of* u *will be found, which will converge to a minimal value of* u*. If our function* u*, assumed not to take negative values, does take null values, these will always be obtained by the above method,* provided that *the values* x, y, z, … *are suitably chosen.*[8]

According to his last words, Cauchy does not seem to believe that the method always finds a solution; yet, he also seems to hope it: see the excerpt of footnote 4. Anyway a simple picture reveals that the least-squares function in (4)

---

[6]Il est aisé d'en conclure que la valeur $\Theta$ de $u$ déterminée par la formule (2), deviendra inférieure à u, si $\theta$ est suffisamment petit. Si, maintenant, $\theta$ vient à croître, et si, comme nous l'avons supposé, la fonction $f(x, y, z, \ldots)$ est continue, la valeur $\Theta$ de $u$ décroîtra jusqu'à ce qu'elle s'évanouisse, ou du moins jusqu'à ce qu'elle coïncide avec une valeur *minimum*, déterminée par l'équation à une seule inconnue (3).

[7]Here we have an additional proposal: least squares, which is some 50 years old. Incidentally, its paternity provoked a dispute between Legendre and Gauss (who peremptorily concluded: *I did not imagine that Mr Legendre could feel so strongly about such a simple idea; one should rather wonder that nobody had it 100 years earlier*).

[8]Si la nouvelle valeur de $u$ n'est pas un *minimum*, on pourra en déduire, en opérant toujours de la même manière, une troisième valeur plus petite encore ; et, en continuant ainsi, on trouvera successivement des valeurs de u[sic] de plus en plus petites, qui convergeront vers une valeur minimum de u[sic]. Si la fonction $u$, qui est supposée ne point admettre de valeurs négatives, offre des valeurs nulles, elles pourront toujours être déterminées par la méthode précédente, pouru que l'on choisisse convenablement les valeurs de x, y, z, ….

may display positive local minima, playing the role of "parasitic" solutions. On the other hand, he seems convinced that, being decreasing, the sequence of $u$-values has to converge to a (local) minimum, or at least a stationary point.

Thus, the above excerpt is fairly interesting, coming from a mathematician among the most rigorous of his century. Admittedly, Cauchy has not given deep thought to the problem: *I'll restrict myself here to outlining the principles underlying* [*my method*], *with the intention to come again over the same subject, in a paper to follow.*[9] However, the *"paper to follow"* does not seem to exist. Let us bet that he has underestimated the difficulty and eventually not been able to crack this tough nut. In fact, we are now aware that some form of *uniformity* is required from the objective's continuity – not mentioning the choice of a *"small enough"* $\theta$, which is also delicate.

## References

[1] A. Cauchy. Méthode générale pour la résolution des systèmes d'équations simultanées. *C. R. Acad. Sci. Paris*, 25:536–538, 1847.

Claude Lemaréchal
INRIA
655 avenue de l'Europe
Montbonnot
38334 Saint Ismier
France
claude.lemarechal@inria.fr

[9] Je me bornerai pour l'instant à indiquer les principes sur lesquels elle se fonde, me proposant de revenir avec plus de détails sur le même sujet, dans un prochain Mémoire.