How to Cite:
Shaker, A.- shimari N., & Al-Jilawi, A. S. (2022). Coordinate descent method for solving constraint optimization problems . *International Journal of Health Sciences*, *6*(S2), 2395–2400. https://doi.org/10.53730/ijhs.v6nS2.5585

Coordinate descent method for solving constraint optimization problems

Al-shimari Nofal Shaker
College of Education for Pure S.C/University of Babylon
Email: nawfal.kazem@student.uobabylon.edu.iq

Ahmed Sabah. Al-Jilawi
College of Education for Pure S.C/University of Babylon
Email: aljelawy2000@yahoo.com

***Abstract*---**The study generally aims to find line search technique with exact line search without relying on update or invertible Hessian matrix at every iteration, and theoretical improvement for if H∈ $R^{n\times n}$ be a symmetric matrix then H has n mutually orthogonal eigenvectors.

***Keywords*---**optimization, line search technique, characteristics Hessian matrix.

Introduction

Line search technique is on the off chance that it looks for the base of a descent direction vector, when processed iteratively with a sensible advance size [1 ,20]. The solution methods for unconstrained optimization problems can be broadly classified into gradient-based and non-gradient based search methods. Optimization theory and techniques are new topics in applied mathematics, operations research, and computational mathematics with a large range of applications in scientific and engineering, business administration, and space technology [2, 5, 16]. This topic participates in the optimum solution of problems that are determined mathematically [6,22], that is, in light of a practical issue, through many schemes and using scientific methods and tools, a better solution to the problem can be obtained. In the late 1940's, optimization became a separate topic when George Bernard Dantzig [7,17] introduced the popular simplex algorithm for linear programming [15,21] .

Coordinate Descent Method

Consider the problem $\min_x f(x)$ where $f: R^n \to R$, where $f \in C^1$ is the first order continuously differential for every coordinate variable x_i, $i = 1, \ldots, n$ minimize $f(x)$

Manuscript submitted: 27 Jan 2022, Manuscript revised: 18 Feb 2022, Accepted for publication: 09 March 2022

with respect to x_i, keeping the other coordinate variable x_j, $j \neq i$ constant[8,18 ,19]. Algorithm 2 : Coordinate Descent Method

1. Initialize x^o,ϵ; set k:=0
2. while $\|g^k\| > \epsilon$
   (a) for $i = 1, \ldots, n$
   (b) $x_i{}^{new} = arg\min_{x_i} f(x)$
   (c) $x_i = x_i{}^{new}$ end while

output: $x^* = x^k$ a stationary point of $f(x)$ globally convergent method if a search along any coordinate direction yields a unique minimum point.

Example 2.1 *Consider the problem the* $\min_x f(x) = 4{x_1}^2 + {x_2}^2 - 2x_1x_2$, *use coordinate descent method with exact line search to solve this problem* $x^o = (-1,-1)^\top$, *let* $d^o = (1,0)^\top$ *and* $x^1 = x^o + \alpha^o d^o$ *where* $\alpha^o = argmin_\alpha \phi(\alpha)$

$$\phi_o(\alpha) = f(x^0 + \alpha^0 d^o)\colon \alpha^0(> 0)$$

$$\phi_o(\alpha) = f\begin{pmatrix} x^o + \alpha^0 {d_1}^o \\ x^o + \alpha^0 {d_2}^o \end{pmatrix} = f\begin{pmatrix} -1 + \alpha^0(1) \\ -1 + \alpha^0(0) \end{pmatrix} = f\begin{pmatrix} -1 + \alpha^0 \\ -1 \end{pmatrix} = 4(\alpha - 1)^2 + 1 - 2(-1)(-1 + \alpha) = 4(\alpha - 1)^2 + 1 + 2(\alpha - 1)\alpha_o{}'(\alpha) = 0 \Rightarrow 8(\alpha - 1) + 2 = 0 \Rightarrow 4(\alpha - 1) + 1 = 0 \Rightarrow \alpha - 1 = \frac{-1}{4} \Rightarrow = \frac{3}{4} \therefore x^1 = (-1,-1) + \frac{3}{4}(1,0)^\top = (-\frac{1}{4}, -1)^\top \Rightarrow d^1 = (0,1)^\top; x^2 = x^1 + \alpha^1 d^1 x^2 = x^1 + \alpha^1 d^1, \alpha^1 = arg\min_\alpha \phi_1(\alpha) \phi_o(x^2) = f\begin{pmatrix} \frac{-1}{4} \\ -1 + \alpha \end{pmatrix} = (\alpha - 1)^2 + \frac{\alpha - 1}{2} + \frac{1}{4}\phi_1{}' = 0 \Rightarrow \alpha^1 = \frac{3}{4} x^2 = x^1 + \alpha^1 d^1 = (-\frac{1}{4}, -\frac{1}{4}) \neq x^*$$

So need some more iteration to reach x^*, the Hessian matrix denote by

$$H = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}$$

since matrix not diagonal and x_1 , x_2 are not separable, then could not be attained in two steps using coordinate descent method (if x^o is not on one of the principal axes of the elliptical contours)

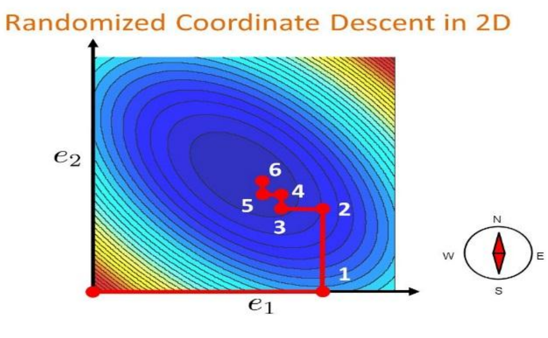

Figure 1: coordinate descent method

Example 2.2 *Consider the problem the* $min_x f(x) = 4{x_1}^2 + {x_2}^2$, *use coordinate descent method with exact line search to solve this problem* $x^o = (-1,-1)^\top$, *let* $d^o = (1{,}0)^\top$ *and* $x^1 = x^o + \alpha^o d^o$ *where* $\alpha^o = argmin_\alpha \phi_o(\alpha) = f(x^O + \alpha d^o)$: $\alpha(> 0)$

$\phi_o(\alpha) = f\begin{pmatrix} x^o + \alpha {d_1}^o \\ x^o + \alpha {d_2}^o \end{pmatrix} = f\begin{pmatrix} -1 + \alpha(1) \\ -1 + \alpha(0) \end{pmatrix} = f\begin{pmatrix} -1 + \alpha \\ -1 \end{pmatrix} = f\begin{pmatrix} -1 + \alpha \\ -1 \end{pmatrix} = 4(\alpha - 1)^2 + 1 = 4(\alpha -)^2 + 1 \Rightarrow {\alpha_o}'(\alpha) = 0 \Rightarrow 8(\alpha - 1) = 0 \Rightarrow 8(\alpha - 1) = 0\alpha - 1 = 0 \therefore x^1 = (1,-1) + (-1{,}0)^\top = (0{,}1) \Rightarrow d^1 = (0{,}1)^\top; x^2 = x^1 + \alpha^1 d^1 x^2 = x^1 + \alpha^1 d^1, \alpha^1 = arg\min_\alpha \phi_1(\alpha) x^2 = x^1 + \alpha^1 d^1 = (0{,}0) = x^*$ since matrix diagonal and x_1 , x_2 are separable, then could be attained in two steps using coordinate descent method$H = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}$

Definition 2.1 *Let* $H \in R^{n\times n}$ *be a symmetric matrix, the vectors* $\{d^o, \dots, d^{n-1}\}$ *are said to be* $H -$ *conjugate if the are linearly independent and* ${d^i}^\top H d^j = 0, \forall i \neq j$*[?]*

Theorem 2.1 A convex quadratic function can be minimized in at most n-steps, provided search along conjugate direction of the Hessian matrix[11].

Proof. consider the problem $\min_x f(x) = \frac{1}{2} x^\top H x + c^\top x$ where H is a symmetric positive definite matrix (H is not diagonal matrix), let$\{d^o, \dots, d^{n-1}\}$ be a set of linearly independent direction and $x^o \in R^n$ any $x \in R^n$ can be represented as $x = x^o + \sum_{i=0}^{n-1} \alpha^i d^i$ then rewrite problem as given $\{d^o, \dots, d^{n-1}\}$ and $x^o \in R^n$, the given problem is to minimize $\psi(\alpha)$ defined as

$$\frac{1}{2}(x^o + \sum_{i=0}^{n-1} \alpha^i d^i)^\top H (x^o + \sum_{i=0}^{n-1} \alpha^i d^i) + c^\top (x^o + \sum_{i=0}^{n-1} \alpha^i d^i)^\top$$

Let us define $D = (d^o | d^1 | \dots | d^{n-1})$ be a matrix denote by $d^o | d^1 | \dots | d^{n-1}$ and $\alpha = (\alpha^o, \dots, \alpha_{n-1})$ $\psi(\alpha)$ in this compact notation

$$\frac{1}{2}\alpha^\top D^\top H D \alpha + (Hx^o + c)^\top D_\alpha + \frac{1}{2}{x^o}^\top H x^o + c^\top x^o$$

Let $K = \frac{1}{2}{x^o}^\top H x^o + c^\top x^o$ where K is constant H^o, x^o and $c^\top$ are constant, then when $\min \psi(\alpha)$ can ignore K then

$$\min \psi(\alpha) = \frac{1}{2}\alpha^\top D H D \alpha + (Hx^o + c)^\top D\alpha$$

the Hessian matrix in this quadratic function.

$$Q = D^\top H D = \begin{pmatrix} {d^o}^\top H d^o & {d^o}^\top H d^1 & \cdots & {d^o}^\top H d^{n-1} \\ {d^1}^\top H d^o & {d^1}^\top H d^1 & \cdots & {d^o}^\top H d^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ {d^o}^\top H d^o & {d^{n-1}}^\top H d^1 & \cdots & {d^{n-1}}^\top H d^{n-1} \end{pmatrix}$$

And the diagonal of $Q = \left({d^o}^\top H d^o, \dots, {d^{n-1}}^\top H d^{n-1}\right)$. Decides to make the Hessian matrix Q is diagonal matrix, the way to do that is that to make all this half diagonal entries in this matrix equal to zero, in other words when ever $i \neq j$ then ${d^i}^\top Hd = 0, \forall i \neq j$

$$Q = D^\top HD = \begin{pmatrix} {d^o}^\top Hd^o & 0 & \cdots & 0 \\ 0 & {d^1}^\top Hd^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {d^{n-1}}^\top Hd^{n-1} \end{pmatrix}$$

this matrix it is easy to inveritable
Therefore

$${Q^{-1}}_{ij} = \begin{cases} \dfrac{1}{d^\top Hd} & \text{if} \quad i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\min\psi(\alpha) = \frac{1}{2}\alpha^\top DHD\alpha + (Hx^o + c)^\top D\alpha + c^\top(x^o + \sum_i \alpha^i d^i)$$

since $c^\top(x^o + \sum_i \alpha^i d^i$ constant and $\psi(\alpha)$ is separable in terms of $\alpha^o, \dots, \alpha^{n-1}$ because this separable in terms it is easy to optimize this objective function individually in terms of α, will optimize the problem with respect to α and in the α space you can think of it is a coordinate descent method, take α^1 at a time and optimize with respect α

$$\frac{\partial\psi}{\partial\alpha^i} = 0 \Rightarrow {\alpha^i}^* = -\frac{d^i(Hx^o + c)}{{d^i}^\top Hd^i}$$

Therefore

$$x^* = x^o + \sum\nolimits_{i=0}^{n-1} {\alpha^i}^* d^i x^* = x^o + (\sum\nolimits_{i=0}^{n-1} -\frac{d^i(Hx^o + c)}{{d^i}^\top Hd^i}) d^i$$

Numerical application

Consider the unconstrained optimization problem.

$$\min_{x \in R^2} f(x) = [1.5 - x_1(1 - x_2)]^2 + [2.625 - x_1(1 - {x_2}^3)]^2 + [2.25 - x_1(1 - {x_2}^2)]^2$$

will applying the different technique of line search by using the python programming to the same problem and will comparing the beaver of converge to the minimal point and explain by study result who is better technique. the date using to solve example are. the current point is $x = [1, -3]$and with the stop condition$\epsilon = 1e - 6$ the fowling output on the python console

The Conjugate Gradient, In this method depend on orthogonal eigenvectors of Hessian matrix, the table(see 1) the behaver of converge to minimizer.

k	x_1	x_2	$f(x_1, x_2)$	$\parallel gradient \parallel$
1	0.157	-2.213	11.263	2200.599
2	2.253	0.041	0.638	0.616
3	2.457	0.325	0.180	2.878
4	2.632	0.412	0.134	1.003
5	3.084	0.508	0.048	1.791
6	3.120	0.499	0.040	0.962
7	3.120	0.497	0.040	0.077
8	3.101	0.489	0.040	0.036
9	3.021	0.471	0.038	0.180
10	3.018	0.472	0.038	0.059
11	3.018	0.472	0.038	0.006

Table 1: Conjugate Gradient Method

Conclusion

The goal of the study focused on the following points:

1. In this method we have not need to invertible or update of Hessian at every iteration.
2. Every symmetric matrix have n (H-conjugate vectors).
3. Object function have n variable, we can proved one of them to minimize object function.
4. In this method we applied the exact line search only.

References

1. Ali Behnood, Emadaldin Mohammadi Golafshani, and Seyedeh Mohaddeseh Hosseini. Determinants of the infection rate of the covid-19 in the us using anfis and virus optimization algorithm (voa). Chaos, Solitons & Fractals, 139:110051, 2020.
2. Mohamed Issa and Mohamed Abd Elaziz. Analyzing covid-19 virus based on enhanced fragmented biological local aligner using improved ions motion optimization algorithm. Applied soft computing, 96:106683, 2020.
3. Pramath Kakodkar, Nagham Kaka, and MN Baig. A comprehensive literature review on the clinical presentation, and management of the pandemic coronavirus disease 2019 (covid-19). Cureus, 12(4), 2020.
4. Swapna rekha Hanumanthu. Role of intelligent computing in covid-19 prognosis: A state-of-the-art review. Chaos, Solitons & Fractals, page 109947, 2020.
5. Kenneth Joseph Arrow, Floyd J Gould, and Stephen Mills Howe. A general saddle point result for constrained optimizations. Technical report, North Carolina State University. Dept. of Statistics,1971.
6. Kenneth J Arrow and Leonid Hurwicz. Reduction of constrained maxima to saddle-point problems. In Traces and Emergence of Nonlinear Programming, pages 61–80. Springer, 2014.
7. Dimitri P Bertsekas and Athena Scientific. Convex optimization algorithms. Athena Scientific Belmont, 2015.

8. Dimitri P Bertsekas. Constrained optimization and Lagrange multiplier methods. Academic press, 2014.
9. Samuel Burer, Renato DC Monteiro, and Yin Zhang. Rank-two relaxation heuristics for max-cut and other binary quadratic programs. SIAM Journal on Optimization, 12(2):503–521,2002.
10. Alice Chiche and Jean Charles Gilbert. How the augmented lagrangian algorithm can deal with an infeasible convex quadratic optimization problem. Journal of Convex Analysis (to appear).[pdf], 3(4):5, 2014.
11. Andrew R Conn, Nick Gould, Annick Sartenaer, and Ph L Toint. Convergence properties of an augmented lagrangian algorithm for optimization with a combination of general equality and linear constraints. SIAM Journal on Optimization, 6(3):674–703, 1996.
12. Stephan Dempe and Jonathan F Bard. Bundle trust-region algorithm for bilinear bilevel programming. Journal of Optimization Theory and Applications, 110(2):265–288, 2001.
13. P Dimitri Bertsekas. Nonlinear programming. 1999. Athena, Scientific, Belmont CA.
14. Mahmoud Mahmoud El-Alem. A global convergence theory for a class of trust region algorithms for constrained optimization. PhD thesis, 1988.
15. Stefan Feltenmark and Krzysztof C Kiwiel. Dual applications of proximal bundle methods, including lagrangian relaxation of nonconvex problems. SIAM Journal on Optimization,10(3):697–721, 2000.
16. Al-Jilawi, A. S., & Abd Alsharify, F. H. (2022). Review of Mathematical Modelling Techniques with Applications in Biosciences. Iraqi Journal For Computer Science and Mathematics, 3(1), 135-144.
17. Alridha, A., Wahbi, F. A., & Kadhim, M. K. (2021). Training analysis of optimization models in machine learning. International Journal of Nonlinear Analysis and Applications, 12(2), 1453-1461.
18. Kadhim, M. K., Wahbi, F. A., & Hasan Alridha, A. (2022). Mathematical optimization modeling for estimating the incidence of clinical diseases. International Journal of Nonlinear Analysis and Applications, 13(1), 185-195.
19. Alridha, A., Salman, A. M., & Al-Jilawi, A. S. (2021, March). The Applications of NP-hardness optimizations problem. In Journal of Physics: Conference Series (Vol. 1818, No. 1, p. 012179). IOP Publishing.
20. Salman, A. M., Alridha, A., & Hussain, A. H. (2021, March). Some Topics on Convex Optimization. In Journal of Physics: Conference Series (Vol. 1818, No. 1, p. 012171). IOP Publishing.
21. Alridha, A., & Al-Jilawi, A. S. (2021, March). Mathematical Programming Computational for Solving NP-Hardness Problem. In Journal of Physics: Conference Series (Vol. 1818, No. 1, p. 012137). IOP Publishing.
22. Alridha, A. H., & Al-Jilawi, A. S. (2022). Solving NP-hard problem using a new relaxation of approximate methods. International Journal of Health Sciences, 6(S3), 523–536. https://doi.org/10.53730/ijhs.v6nS3.5375