Maltepe Journal of Mathematics
ISSN:2667-7660, URL: https://dergipark.org.tr/tr/pub/mjm
Volume I Issue 2 (2019), Pages 89-95.

ON CONVEX OPTIMIZATION IN HILBERT SPACES

BENARD OKELO

WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER, MATHEMATISCHES INSTITUT
EINSTEINSTR. 62, 48149-MÜNSTER, GERMANY, ORCID: 0000-0003-3963-1910

Abstract. In this paper, convex optimization techniques are employed for convex optimization problems in infinite dimensional Hilbert spaces. A first order optimality condition is given. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $x \in \mathbb{R}^n$ be a local solution to the problem $\min_{x\in\mathbb{R}^n} f(x)$. Then $f'(x, d) \geq 0$ for every direction $d \in \mathbb{R}^n$ for which $f'(x, d)$ exists. Moreover, Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $x^* \in \mathbb{R}^n$. If x^* is a local minimum of f, then $\nabla f(x^*) = 0$. A simple application involving the Dirichlet problem is also given. Lastly, we have given optimization conditions involving positive semi-definite matrices.

1. INTRODUCTION

Studies on convex optimization have been carried out by many mathematicians and it still remains interesting. Convex operators, convex vector-functions among others, that is, mappings defined on a convex subset of a vector space and with values in an ordered vector space, have been intensively studied in the last years, mainly in connection with optimization problems and mathematical programming in ordered vector spaces (see [1], [3], [5]). The normality of the cone is essential in the proofs of the continuity properties of convex vector-functions. Lipschitz properties of continuous convex vector functions defined on an open convex subset of a normed space and with values in a normed space ordered by a normal cone have also been considered [6]. Equicontinuity results for pointwise bounded families of continuous convex mappings have also been studied with many interesting results obtained. It has been shown that a pointwise bounded family of continuous convex mappings, defined on an open convex subset of a Banach space X and with values in a normed space Y ordered by a normal cone, is locally equi-Lipschitz on X. Equicontinuity and equi-Lipschitz results for families of continuous convex mappings defined on open convex subsets of Baire topological vector spaces or of barrelled locally convex spaces and taking values in a topological vector space respectively in a locally convex space, ordered by a normal cone have also been obtained [7]. We are concerned here with the classical results on optimization of convex functionals in infinite-dimensional real Hilbert spaces. When working with infinite-dimensional spaces, a basic difficulty is that, unlike the case in finite-dimension, being closed and bounded

2010 *Mathematics Subject Classification.* Primary: 49M30 Secondaries: 65K10; 90C26; 93B40.
Key words and phrases. Optimization; Convexity; Hilbert space.

does not imply that a set is compact. In reflexive Banach spaces, this problem is mitigated by working in weak topologies and using the result that the closed unit ball is weakly compact. This in turn enables mimicking some of the same ideas in finite-dimensional spaces when working on unconstrained optimization problems. It is the goal of these note to provide a concise coverage of the problem of minimization of a convex function on a Hilbert space ([8]-[10]). The focus is on real Hilbert spaces, where there is further structure that makes some of the arguments simpler. Namely, proving that a closed and convex set is also weakly sequentially closed can be done with an elementary argument, whereas to get the same result in a general Banach space we need to invoke Mazur's Theorem. The ideas discussed in this brief note are of great utility in theory of Partial Differential Equations, where weak solutions of problems are sought in appropriate Sobolev's spaces [2]. After a brief review of the requisite preliminaries, we develop the main results. Though, the results in this note are classical, we provide proofs of key theorems for a self contained presentation. A simple application, regarding the Dirichlet problem, is provided for the purposes of illustration. Also, we recall an important point about notions of compactness and sequential compactness in weak topologies [4]. It is common knowledge that compactness and sequential compactness are equivalent in metric spaces. The situation is not obvious in the case of weak topology of an infinite-dimensional normed linear space [6]. Lastly, we give optimization conditions involving positive semi-definite matrices.

2. PRELIMINARIES

Definition 2.1. *A sequence x_n in a Banach space B is said to converge to $x \in B$ if $\lim_{n\to\infty} x_n = x$. Also a sequence x_n in a Hilbert space H converges weakly to x if, $\lim_{n\to\infty}\langle x_n, u\rangle = \langle x, u\rangle$, $\forall u \in H$. We use the notation $x_n \rightharpoonup x$ to mean that x_n converges weakly to x.*

Definition 2.2. *A set $D \subseteq \mathbb{R}^n$ is bounded if there exists a constant $M > 0$ such that $\|x\| < M$, for all $x \in D$. The set Dis said to be compact if it is closed and bounded.*

Example 2.1. *A closed interval $[a, b]$ is bounded in $\mathbb{R}$, and is therefore also compact. The circle and its interior $\{(x, y)|x^2 + y^2 \leq 1\}$ is a closed set in $\mathbb{R}^2$, and is also bounded, and therefore it is compact. The interval $[0,\infty)$ is closed in $\mathbb{R}$, as its complement $(-\infty, 0)$ is open, but it is not bounded, so it is not compact either.*

Definition 2.3. *A real valued function f on a Banach space B is lower semi-continuous (LSC) if $f(x) \leq \liminf_{n\to\infty} f(x_n)$ for all sequences x_n in B such that $x_n \to x$ (strongly) and weakly sequentially lower semi-continuous (weakly sequentially LSC) if $x_n \rightharpoonup x$.*

Definition 2.4. *A non-empty set W is said to be convex if for all $\beta \in [0, 1]$ and $\forall\, x, y \in W$ $\beta x + (1 - \beta)y \in W$. Let X be a metric space and $W \subseteq X$ a non-empty convex set. A function $f : W \to \mathbb{R}$ is convex if for all $\beta \in [0, 1]$ and $\forall\, x, y \in W$*

$$f(\beta x + (1 - \beta)y) \leq \beta f(x) + (1 - \beta)f(y).$$

Remark. *We note that the function f in the above definition is called strictly convex if the above inequality is strict for $x \neq y$ and $\beta \in (0, 1)$. A function f is convex if and only if its epigraph, epi(f), is convex whereby epi$(f) := f(x, r) \in$*

$dom(f) \times \mathbb{R} : f(x) \leq r$. An optimization problem is convex if both the objective function and feasible set are convex.

Definition 2.5. *Let $\mathbb{R}^n$ be an n-dimensional real space and $W \subseteq \mathbb{R}^n$. A point $x^* \in \mathbb{R}^n$ is called a* global minimizer *of the optimization problem* $\min_{x \in W} f(x)$, *if $x^* \in W$ and $f(x^*) \leq f(x)$, for all $x \in W$.*

Definition 2.6. *Let $\mathbb{R}^n$ be an n-dimensional real space and $W \subseteq \mathbb{R}^n$. A point $x^* \in \mathbb{R}^n$ is called a* local minimizer *of the optimization problem* $\min_{x \in W} f(x)$, *if there exists a neighbourhood N of x^* such that x^* is a global minimizer of the problem $\mathcal{P} = \min_{x \in W \cap N} f(x)$. That is there exists $\varepsilon > 0$ such that $f(x^*) \leq f(x)$, whenever $x^* \in W$ satisfies $\|x^* - x\| \leq \varepsilon$.*

Remark. *Any local minimizer of a convex optimization problem is a global minimizer.*

Theorem 2.1. *(Weierstrass Extreme Value Theorem) Every continuous function on a compact set attains its extreme values on that set.*

Proposition 2.2. *Let B be a Banach space and $f : B \to \mathbb{R}$. Then the following are equivalent. (i). f is (weakly sequentially) LSC.*
(ii). $epi(f)$, is (weakly sequentially) closed.

Remark. *$f : B \to \mathbb{R}$ is coercive if for all $x \in B$, $\lim_{\|x\| \to \infty} f(x) = \infty$. As an example, the function $f(x, y) = x^2 + y^2$ is coercive, as $\lim_{\|x\| \to \infty} f(x, y) = \lim_{\|x\| \to \infty} \|x\|^2 + \infty$. Also, A linear function is never coercive. For instance, a linear function on $\mathbb{R}^2$ has the form $f(x, y) = ax + by + c$, for constants a, b and c, and is equal to c along the line defined by the equation $ax + by = 0$. Since $\|x\| \to \infty$ along this line, but $f(x, y) = c$ along this line, $f(x, y)$ is not coercive. As these examples show, in order for a function to be coercive, it must approach $+\infty$ along any path within $\mathbb{R}^n$ on which $\|x\|$ becomes infinite.*

Proposition 2.3. *Let $f(x)$ be a continuous function defined on all of $\mathbb{R}^n$. If $f(x)$ is coercive, then $f(x)$ has a global minimizer. Furthermore, if the first partial derivatives of $f(x)$ exist on all of $\mathbb{R}^n$, then any global minimizers of $f(x)$ can be found among the critical points of $f(x)$.*

Lemma 2.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous on all of $\mathbb{R}^n$. The function f is coercive if and only if for every $\beta \in \mathbb{R}$ the set $\{x | f(x) \leq \beta\}$ is compact.*

Proof. First we need to show that the coercivity of f implies the compactness of the sets $\{x | f(x) \leq \beta\}$. We note that the continuity of f implies the closedness of the sets $\{x | f(x) \leq \beta\}$. Therefore, it suffices to show that any set of the form $\{x | f(x) \leq \beta\}$ is bounded. We prove this by contradiction. Suppose to the contrary that there is an $\beta \in \mathbb{R}$ such that the set $S = \{x | f(x) \leq \beta\}$ is unbounded. Then there must exist a sequence $\{x^r\} \subset S$ with $\|x^r\| \to \infty$. But then, by the coercivity of f, we must also have $f(x^r) \to \infty$. This contradicts the fact that $f(x^r) \leq \beta$ for all $r = 1, 2, ...$ Hence the set S must be bounded. Conversely, assume that that each of the sets $\{x | f(x) \leq \beta\}$ is bounded and let $\{x^r\} \subset \mathbb{R}$ be such that $\|x^r\| \to \infty$. Assume that there exists a subsequence of the integers $J \subset \mathbb{N}$ such that the set $\{f(x^r)\}_J$ is bounded above. Then there exists $\beta \in \mathbb{R}$ such that $\{f(x^r)\}_J \subset \{x | f(x) \leq \beta\}$. But this cannot be the case since each of the sets $\{x | f(x) \leq \beta\}$ is bounded while every subsequence of the sequence $\{x^r\}$ is unbounded by definition. Therefore, the set $\{f(x^r)\}_J$ cannot be bounded, and so the sequence $\{f(x^r)\}$ contains no bounded subsequence, that is $f(x^r) \to \infty$. □

Corollary 2.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous on all of $\mathbb{R}^n$. If f is coercive, then f has at least one global minimizer.*

Proof. Let $\beta \in \mathbb{R}$ be chosen so that the set $S = \{x|f(x) \leq \beta\}$ is non-empty. By coercivity, this set is compact. By Weierstrass's Theorem, the problem $\min\{f(x)|x \in S\}$ has at least one global solution. It is easy to see that the set of global solutions to the problem $\min\{f(x)|x \in S\}$ is a global solution to $\mathcal{P}$ and this completes the proof. □

Remark. *We note that coercivity hypothesis is stronger than as strictly required in order to establish the existence of a solution. Indeed, a global minimizer must exist if there exist one non-empty compact lower level set. We do not need all of them to be compact. However, in practice, coercivity is a sufficiency.*

Proposition 2.6. *Let H be an infinite dimensional real separable Hilbert space and let $W \subseteq H$ be a (strongly) closed and convex set. Then, W is weakly sequentially closed.*

Proof. Let the sequence $x_n \rightharpoonup x$ be in W. It only suffices to show that $x \in W$ by showing that $x = \phi_W(x)$, where $\phi_W(x)$ is the projection of x into the closed convex set W. Indeed, we know that the projection $\phi_W(x)$ satisfies the variational inequality, $\langle x - \phi_W(x), y - \phi_W(x)\rangle \leq 0$, for all $y \in W$.
So,

$$\langle x - \phi_W(x), x_n - \phi_W(x)\rangle \leq 0,\ \forall\, n. \tag{2.1}$$

But, $x_n \rightharpoonup x$ be in W so we have,

$$\begin{aligned} \|x - \phi_W(x)\|^2 &= \langle x - \phi_W(x), x - \phi_W(x)\rangle \\ &= \lim_{n\to\infty} \langle x - \phi_W(x), x_n - \phi_W(x)\rangle \end{aligned}$$

Hence, by Equation 2.1 we have $\|x - \phi_W(x)\| = 0$. That is, $x = \phi_W(x)$. □

Lemma 2.7. *Let $f : H \to \mathbb{R}$ be a LSC convex function. Then f is weakly LSC.*

Proof. We know that f is convex iff $epi(f)$ is convex. Moreover, $epi(f)$ is strongly closed because f is (strongly) LSC. By proposition 2.6 we have that $epi(f)$ is weakly sequentially closed implying that f is weakly sequentially LSC.

□

3. MAIN RESULT

Theorem 3.1. *Let H be an infinite dimensional real separable Hilbert space and $W \subseteq H$ be a weakly sequentially closed and bounded set. Let $f : W \to \mathbb{R}$ be weakly sequentially LSC. Then f is bounded from below and has a minimizer on W.*

Proof. The proof has two steps:
(i). f is bounded below.
(ii). There exists a minimizer in W.

Step(i): Suppose that f is not bounded from below. Then there exist a sequence $x_n \in W$ such that $f(x_n) < -n$ for all n. But W is bounded so x_n has a

weakly convergent subsequence x_{n_i} Furthermore, W is weakly sequentially closed therefore $x \in W$. Then, since f is weakly sequentially LSC we have $f(x) \leq \liminf_{n\to\infty} f(x_{n_i}) = -\infty$ which is a contradiction. Hence, f is bounded from below.
Step(ii): Let $x_n \in W$ be a minimizing sequence for f that is $f(x_n) \to \inf_W f(x)$. Let $\lambda := \inf_W f(x)$. Since W is bounded and weakly sequentially closed, it follows that x_n has a weakly convergent subsequence has a weakly convergent subsequence $x_{n_i} \in W$. But f is weakly sequentially LSC so we have

$$\lambda \leq f(x^*) \leq \liminf f(x_{n_i}) = \lim f(x_{n_i}) = \lambda$$

So, $f(x^*) = \lambda$ □

Corollary 3.2. *Let H be an infinite dimensional real separable Hilbert space and $W \subseteq H$ be a weakly sequentially closed and bounded set. Let $f : W \to \mathbb{R}^n$ be non-empty and closed, and that $f : W \to \mathbb{R}^n$ is LSC and coercive. Then the optimization problem $\inf_{x\in W} f(x)$ admits at least one global minimizer.*

Proof. With an analogy to the proof of Theorem 3.1 the proof of coercivity is sufficient. □

Theorem 3.3. *A function that is strictly convex on W has a unique minimizer on W.*

Proof. Assume the contrary, that $f(x)$ is convex yet there are two points $x, y \in W$ such that $f(x)$ and $f(y)$ are local minima. Because of the convexity of W every point on the secant line $\beta x + (1-\beta)y$ is in W. Without loss of generality suppose $f(x) \geq f(y)$ if this is not the case, simply relabel the points. We then have $\beta f(x) + (1-\beta)f(y) < f(y), \forall\ \beta \in (0,1)$. But f is strictly convex, we also have $f(\beta x + (1-\beta)y) < f(x), \forall\ \beta \in (0,1)$. Taking β arbitrarily close to 0 along the secant line, $z = \beta x + (1-\beta)y$ remains in W(since W is convex) and $f(z)$ remains strictly below $f(x)$ (because f is strictly convex). Therefore, there is no open ball B containing x such that $f(x) < f(z), \forall\ z(B \cap W) \setminus x$. Therefore, x is not a local minimizer, which is a contradiction. □

In this last part we give an optimality conditions. We give the first order condition for optimality here. Consider the function $\psi : \mathbb{R} \to \mathbb{R}$ given by $\psi(t) = f(x+td)$ for some choice of x and d in $\mathbb{R}^n$. The key variational object in this context is the directional derivative of f at a point x in the direction d given by

$$f'(x,d) = \lim_{t\downarrow 0} \frac{f(x+td) - f(x)}{t}.$$

When f is differentiable at the point $x \in \mathbb{R}^n$, then $f'(x,d) = \nabla f(x)^T d = \psi'(0)$. The next two results give us an optimality condition.

Proposition 3.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $x \in \mathbb{R}^n$ be a local solution to the problem $\min_{x\in\mathbb{R}^n} f(x)$. Then $f'(x,d) \geq 0$ for every direction $d \in \mathbb{R}^n$ for which $f'(x,d)$ exists.*

Theorem 3.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $x^* \in \mathbb{R}^n$. If x^* is a local minimum of f, then $\nabla f(x^*) = 0$.*

Proof. We know that every differentiable function is continuous so by Proposition 3.4 we have we have

$$0 \leq f'(x^*, d) = \nabla f(x^*)^T d,$$

for all $d \in \mathbb{R}^n$. Taking $d = -\nabla f(x^*)$ we obtain $0 \leq -\nabla f(x^*)^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 \leq 0$. Therefore, $\nabla f(x^*) = 0$. □

Example 3.1. *Consider the Dirichlet problem:* $-\triangle u = f$, *in* W *and* $u = 0$, *on* ∂W, *where* $W \subset \mathbb{R}^n$ *is a bounded domain, and* $f \in L^2(W)$. *It is well known that this problem has a weak solution which is convex and continuous, and coercive. Thus, the existence of a unique minimizer is ensured by application of Theorem 3.5.*

In the next results we consider positive definite matrices. We use concepts from linear algebra to obtain simpler, more intuitive criteria for determining whether a symmetric matrix, such as the Hessian of a function at a point, is positive or negative definite or semi-definite. Let T be an $n \times n$ symmetric matrix. A nonzero vector $x \in \mathbb{R}^n$ is an eigenvector of T if there exists a scalar λ such that $Tx = \lambda x$. The scalar λ is called an eigenvalue of T corresponding to x. From the equation $Tx - \lambda x = (T - \lambda I)x = 0$, and the fact that $x \neq 0$ it follows that the matrix $T - \lambda I$ is not invertible. Therefore, any eigenvalue λ of T satisfies $det(T - \lambda I) = 0$. This determinant is a polynomial of degree n in λ, which is called the characteristic polynomial. Therefore, the eigenvalues can be found by computing the characteristic polynomial, and then computing its roots. For a general matrix T, the eigenvalues may be real or complex, because a polynomial with real coefficients can have complex roots, but the eigenvalues of a symmetric matrix T are real. Furthermore, if T is symmetric, there exists an orthogonal matrix P, meaning that $P^t P = I$, such that $T = PDP^t$, where D is a diagonal matrix whose diagonal entries are the eigenvalues of T. The columns of P are orthonormal vectors, meaning that they are orthogonal and are of magnitude 1. They are also the eigenvectors of T. The following result follows immediately.

Theorem 3.6. *Let* T *be a symmetric matrix on a real Hilbert space. Then the following conditions hold:*
(i). T *is positive definite if and only if all of its eigenvalues are positive;*
(ii). T *is negative definite if and only if all of its eigenvalues are negative;*
(iii). T *is positive semi-definite if and only if all of its eigenvalues are nonnegative;*
(iv). T *is negative semi-definite if and only if all of its eigenvalues are non-positive;*
(v). T*is indefinite if and only if at least one of its eigenvalues is positive and at least one of its eigenvalues is negative.*

Proof. The proof is trivial. □

Next we demonstrate the use of these conditions for optimization in the next example.

Example 3.2. *Let* $f(x, y, z) = x^2 + y^2 + z^2 - 4xy$. *Then we have* $\nabla f(x, y, z) = (2x - 4y, 2y - 4x, 2z)$, *which yields the critical point* $(0, 0, 0)$, *and*
$Hf(x, y, z) = \begin{pmatrix} 2 & -4 & 0 \\ -4 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. *This matrix has the characteristic polynomial* $detHf(x, y, z) - \lambda I = (2 - \lambda)(\lambda + 2)(\lambda - 6)$. *Therefore, the eigenvalues are* $2, -2$ *and* 6, *which means that the Hessian is indefinite. We conclude that* $(0, 0, 0)$ *is a saddle point, and there are no global maximizers or minimizers.*

4. CONCLUSION

This work is geared to its extension to portfolio optimization, whereby applications to stochastic optimization with regarding Cox-Ross-Rubinstein model and Hamilton-Jacobi-Bellman Equation will be considered.

Acknowledgement. The author is grateful to the School of Mathematics and Actuarial Science, Jaramogi Oginga Odinga University of Science and Technology, Kenya for the permission to carry out postdoctorate research visit at the Institute of Mathematics, University of Muenster, Germany.

References

[1] F. Albiac and N. J. Kalton, *Topics in Banach space theory*, Volume 233 of Graduate Texts in Mathematics. Springer, New York, 2006.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, United Kingdom, 2004.

[3] D. P. Bertsekas, *Convex Analysis and Optimization*, Athena Scienti.c, Belmont, MA, 2003.

[4] N. Dunford and J. T. Schwartz, *Linear operators*, Part I. Wiley Classics Library. John Wiley and Sons Inc., New York, 1988.

[5] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North Holland, Amsterdam, 1976.

[6] I. Ekeland and T. Turnbull, *Infinite Dimensional Optimization and Convexity*, The University of Chicago Press, Chicago, 1983.

[7] R. Glowinski, J. L. Lions and R. Tremolieres, *Numerical Analysis of Variational Inequalities*, North Holland, Amsterdam, 1981.

[8] M. Grasmair, *Minimizers of optimization problems*, To appear.

[9] A.J. Kurdila and M. Zabarankin., *Convex functional analysis*, Systems and Control: Foundations and Applications. Birkhauser Verlag, Basel, 2005.

[10] J.P. Vial, *Strong convexity of set and functions*, J. Math. Econom **9** (1982), 187-205.

Benard Okelo,
Westfälische Wilhelms-Universität Münster, Mathematisches Institut Einsteinstr. 62, 48149-Münster, Germany
E-mail address: okelonya@uni-muenster.de