Convex Functions, Majorization Properties and the Convex Conjugate Transform

Derek Brennan, Department of Mathematics and Statistics,
McGill University, Montreal

May, 2005

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Your file Votre référence
ISBN: 0-494-12406-7
Our file Notre référence
ISBN: 0-494-12406-7

Abstract

Let $K_w(t) := \inf_{\|P\|_2^2 \leq t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$, where $w \geq 0$ is a decreasing function and $P \in L^2([0,1], Leb.)$.
Then we have $K_w \leq K_v \Leftrightarrow \int_0^1 \frac{1}{w(\theta)+\lambda} d\theta \geq \int_0^1 \frac{1}{v(\theta)+\lambda} d\theta \; \forall \, \lambda \geq 0$.
We present two similar proofs of this result, which are analogous to the well-known majorization theorem: Let $v, w \geq 0$ be decreasing functions, and suppose $\int_0^1 w(\theta) d\theta = \int_0^1 v(\theta) d\theta$.
Then $F_w \leq F_v \Leftrightarrow \int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta \; \forall \, x \geq 0$,
where $F_w(t) = \int_t^1 w(\theta) d\theta$. Since our proofs of this result rely mainly on the convex conjugate transform, or Legendre transform, we include an exposition of convex functions and convex conjugate transforms.

Résumé

Soit $K_w(t) := \inf_{||P||_2^2 \leq t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$, où $w \geq 0$ est une fonction décroissante et $P \in L^2([0,1],\text{Leb.})$.
Alors on a $K_w \leq K_v \Leftrightarrow \int_0^1 \frac{1}{w(\theta)+\lambda} d\theta \geq \int_0^1 \frac{1}{v(\theta)+\lambda} d\theta$ pour tout $\lambda \geq 0$.
On présente deux preuves semblabes de ce résultat, analogues au théorème de majoration bien-connu: Soit $v, w \geq 0$ et décroissantes, et supposons que $\int_0^1 w(\theta) d\theta = \int_0^1 v(\theta) d\theta$.
Alors $F_w \leq F_v \Leftrightarrow \int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta$ pour tout $x \geq 0$, où $F_w(t) = \int_t^1 w(\theta) d\theta$. Puisque nos preuves de ce résultat s'appuient essentiellement sur la transformée conjuguée convexe, ou la transformée de Legendre, on inclut une exposition sur les fonctions convexes et la transformée conjuguée convexe.

Acknowledgements

I would like to thank my supervisor, Dr. Ivo Klemes, for his excellent combination of guidance, motivation, and fostering of self-direction, as well as for his financial support.

I would also like to thank the Department of Mathematics and Statistics of McGill University for financial support and teaching assistant appointments.

Contents

1 Introduction

We use the following notation:

$\mathbb{R}$	Real numbers
$\mathbb{C}$	Complex numbers
$\|E\|$	Lebesgue measure of the set E
$\int_0^1 \cdot d\theta$	integration with respect to Lebesgue measure on $[0,1]$
χ_E	characteristic function of the set E
L^p	$L^p([0,1], Leb),\ p=1,2$
$\\|f\\|_p$	$\int_0^1 \|f(\theta)\|^p d\theta,\ p=1,2$

Our purpose is to study the properties of the functional K_w to be defined below. This functional was introduced and its main property asserted in Klemes ([5], Prop. 3.3.1); moreover, its definition was motivated there as follows.

Given two functions $v, w \geq 0$ defined on $[0,1]$, such that the normalization condition $\int_0^1 w(\theta)d\theta = \int_0^1 v(\theta)d\theta$ holds, w is said to majorize v if

$$\int_0^t w(\theta)d\theta \geq \int_0^t v(\theta)d\theta \ \forall\, t \in [0,1]$$

A treatment of majorization results can be found in Marshall and Olkin [6].

We recall the following well-known majorization theorem due to Hardy, Littlewood, and Pólya [4]:

Theorem Let $v, w \geq 0$ be decreasing functions, and let

$$F_w(t) = \int_t^1 w(\theta)d\theta, \quad 0 \leq t \leq 1$$

Suppose that v, w satisfy the normalization condition $\int_0^1 w(\theta)d\theta = \int_0^1 v(\theta)d\theta$. Then we have

$$F_w(t) \leq F_v(t)\ \forall\, t \in [0,1] \Leftrightarrow \int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta\ \forall\, x \geq 0$$

□

Remark: This majorization property is very useful in practice because it lends to proofs of inequalities of the form

$$\int \phi(w(\theta))d\theta \geq \int \phi(v(\theta))d\theta$$

where ϕ is any convex function.
In particular, if we take $\phi(x) = x^p$ with $p \geq 1$, then we obtain p-norm inequalities, which are much studied and used in analysis.

We then generalize to the functional

$$J_w(t) := \inf_{||P||_1 \leq t} \int_0^1 |1 - P(\theta)|w(\theta)d\theta, \quad 0 \leq t \leq 1$$

where $w \geq 0$ (not necessarily decreasing) and $P \in L^1([0,1], Leb.)$.
Up to a decreasing rearrangement of w (see [2]), J_w and F_w are essentially identical. To see this, one can show that the infimum is attained precisely when $P = \chi_{[0,t]}$.

We view J_w as an intermediate functional, serving as the motivation for arriving at the consideration of our target functional:

$$K_w(t) := \inf_{||P||_2^2 \leq t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta, \quad 0 \leq t \leq 1$$

where again, $w \geq 0$ but need not be decreasing.
We shall prove the following result, suggested in ([5], Prop. 3.3.1):

Theorem If $v, w \geq 0$, then

$$K_w(t) \leq K_v(t) \; \forall \, t \in [0,1] \Leftrightarrow \int_0^1 \frac{1}{w(\theta) + \lambda} d\theta \geq \int_0^1 \frac{1}{v(\theta) + \lambda} d\theta \; \forall \, \lambda \geq 0$$

Despite the fact that v, w need not be decreasing and that the normalization condition need not hold in order for the statement of the theorem to be true, we view this theorem as an analogue of the aforementioned majorization result for F_w, and refer to these as 'majorization' results on J_w and K_w.

Since the convex conjugate transform is the crux of the proofs of these results, we devote Section 2 to an exposition of convex functions and convex conjugate transforms, culminating in a proof of its involutivity on convex functions.

In Section 3, we first establish that F_w is a convex function. Then

we proceed to give two proofs of the majorization result for F_w. The common element in both proofs is the convex conjugate transform, or "Legendre Transform" suggested by Klemes ([5], Prop. 3.3.1). The first proof is direct and self-contained, and uses a modified convex conjugate transform applied to F_w, followed by another modified transform that inverts the latter.
For the second proof, we show that our modified transform can be viewed as a regular convex conjugate transform of F_w, and then we show the majorization result by invoking involutivity of the transform.

We briefly introduce the intermediate functional J_w in Section 4, stressing that we mention it mainly as the link between the treatments of the functional F_w of the previous section and the functional K_w of the following section.

In Section 5, we give two proofs that K_w is a convex function, and we generalize the first proof (using the modified transform) to prove the majorization property of K_w.

2 The Convex Conjugate Transform of a Convex Function

2.1 Convex Functions

We follow the treatment of Stoer [7] to show that this transform is an involution.

We consider functions $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$. Notice the inclusion of ∞ as a possible function value. The concept of convexity of a function can be extended to include $-\infty$ as a function value, and the ensuing results can be proved with minor modifications (as in [7]), but for our purposes we will not need to include $-\infty$, since all functions we will deal with will be ≥ 0.

Definition 1 *For any function $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$,*

$$K(f) := \{x | f(x) < \infty\}$$

$K(f)$ is also called the domain of finiteness of f.

Definition 2 *A function $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is called convex if it satisfies the following two properties:*

$$1) K(f) \neq \emptyset \quad \textit{(i.e. } f \textit{ assumes at least one finite value)}$$
$$2) \textit{For any } x, y \in K(f), f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$
$$\forall\, 0 \leq \lambda \leq 1$$

Geometrically, property 2) says that between any two points x, y with $f(x)$ and $f(y)$ finite, the line segment joining $f(x)$ and $f(y)$ lies entirely above the graph of f. Note also that the r.h.s. of 2) is well defined in the sense that it is never "$\infty - \infty$" because of the above restriction that $\forall\, x, f(x) \neq -\infty$.

Definition 3 *A convex function f is called strictly convex if strict inequality holds in 2) for all $\lambda \neq 0, 1$ and $x \neq y$.*

The cases $\lambda = 0$ or 1, and $x = y$ (any λ) were excluded from the definition of strict convexity, since 2) always reduces to equality in those cases.

Examples The function $f(x) := x$ is convex but not strictly convex, while the function $f(x) := x^2$ is strictly convex.

The function f defined by

$$f(x) := \begin{cases} 0, \text{ if} -1 \leq x \leq 1 \\ \infty, \text{ otherwise} \end{cases}$$

is a convex function. Indeed, property 2) is still satisfied because the inequality in 2) need only hold for all $x, y \in K(f)$.

□

Lemma 1 *If f is a convex function, then $K(f)$ is a convex set.*

Proof If $K(f)$ only contains 1 point then it is trivially convex, so assume $K(f)$ contains at least 2 points. Suppose it were not convex. Then $\exists x, y \in K(f)$ and a $\lambda \in (0, 1)$ such that $\lambda x + (1 - \lambda)y \notin K(f)$; i.e. such that $f(\lambda x + (1 - \lambda)y) = \infty$. But $f(x), f(y) < \infty$, $\Rightarrow \lambda f(x) + (1 - \lambda)f(y) < \infty$, contradicting inequality 2) in the definition of convexity of f.

□

Now, a convex set in $\mathbb{R}$ is just an interval, so this lemma says that the domain of finiteness of a convex function is an interval. Geometrically, this means that the graph of a convex function $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$

cannot exhibit an asymptotic discontinuity at a point in the interior of its domain.

2.2 Convex Conjugate Functions: Basic Ideas, Geometric Interpretation, and Involutivity

Definition 4 *For any function $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$, define a function $f^c : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ by*

$$f^c(p) := \sup_{x \in \mathbb{R}}(px - f(x)) = \sup_{x}(px - f(x))$$

f^c is called the convex conjugate of f.

Remark In some literature, the convex conjugate transform is called the Legendre transform (e.g. [1],[5]); the treatment of this transform was developed in large part by Fenchel [7].

As the name suggests, one would expect f^c to always be a convex function. Indeed, inequality 2) in the definition of convexity is always satisfied. To see this, take any $x, y \in \mathbb{R}$, and $\lambda \in [0, 1]$.

We need to show that $f^c(\lambda x + (1-\lambda)y) \leq \lambda f^c(x) + (1-\lambda) f^c(y)$.

$$\begin{aligned}
f^c(\lambda x + (1-\lambda)y) &= \sup_{t\in\mathbb{R}}\{[\lambda x + (1-\lambda)y]t - f(t)\} \\
&= \sup_{t\in\mathbb{R}}\{\lambda xt + (1-\lambda)yt - [\lambda f(t) + (1-\lambda)f(t)]\} \\
&= \sup_{t\in\mathbb{R}}\{\lambda[xt - f(t)] + (1-\lambda)[yt - f(t)]\} \\
&\leq \sup_{t\in\mathbb{R}}\{\lambda[xt - f(t)]\} + \sup_{t\in\mathbb{R}}\{(1-\lambda)[yt - f(t)]\} \\
&= \lambda \sup_{t\in\mathbb{R}}\{xt - f(t)\} + (1-\lambda)\sup_{t\in\mathbb{R}}\{yt - f(t)\} \\
&\quad (\text{as } \lambda, 1-\lambda \geq 0) \\
&= \lambda f^c(x) + (1-\lambda) f^c(y)
\end{aligned}$$

Condition 1) is not always satisfied, but it is when f itself is convex:

Lemma 2 *If f is convex, then f^c is also convex.*

Proof By the above remarks we need only check that condition 1) in the definition of convexity holds. For this we need to check two things:

1) that the set $\{p : f(p) = -\infty\} = \emptyset$, and that $K(f^c) \neq \emptyset$.
Since $K(f) = \{x : f(x) < \infty\} \neq \emptyset$, there exists an $x_0 \in K(f)$; i.e. there is an x_0 such that $f(x_0) < \infty$. But then for any $p \in \mathbb{R}$, we have

$$f^c(p) = \sup_x (px - f(x)) \geq px_0 - f(x_0) > -\infty$$

So the set $\{p : f(p) = -\infty\} = \emptyset$.

For the second part, we need to find a $p \in \mathbb{R}$ such that $f^c(p) = \sup_x (px - f(x)) < \infty$.
If $K(f)$ only contains one point x_0, then $px - f(x) = -\infty$ for all $x \neq x_0$,

whence $f^c(p) = \sup_x (px - f(x)) = px_0 - f(x_0) < \infty$.

So we can assume that $K(f)$ contains at least two points. Take any two such points x_0, x_2 and let $x_0 < x_2$. By convexity of f, the line segment joining the points $(x_0, f(x_0))$ and $(x_2, f(x_2))$ lies entirely above the graph of f. Now, the line L through this segment has slope $p = \frac{f(x_2)-f(x_0)}{x_2-x_0}$, and some y-intercept b, so the equation of the line is $y = px + b$. Moreover, $f^(x) < \infty \forall x$, so we have that

$$
\begin{aligned}
&0 \leq (px+b) - f(x) < \infty \forall x \in [x_0, x_2] \\
\Rightarrow &0 \leq \sup_{x \in [x_0, x_2]} ((px+b) - f(x)) = M < \infty
\end{aligned}
$$

Now, intuitively it is clear that the line L lies below the graph of f, outside $[x_0, x_2]$; i.e. $px + b \leq f(x)\ \forall x \notin [x_0, x_2]$.
Formally, assume first that there were some $\overline{x} < x_0$ with $f(\overline{x}) < L(\overline{x})$. Take any point x_1 with $x_0 < x_1 < x_2$. Then $f(x_1) \leq L(x_1)$ by convexity of f. Now consider the line segment L' connecting $f(\overline{x})$ and $f(x_1)$. Then we have

$$
\begin{aligned}
&L'(\overline{x}) = f(\overline{x}) < L(\overline{x}) \text{ by assumption} \\
\text{and } &L'(x_1) = f(x_1) \leq L(x_1) \ (= \text{by assumption}, \leq \text{by convexity of } f)
\end{aligned}
$$

But this means that $L'(x) < L(x)\ \forall x \in [\overline{x}, x_1]$. In particular, $L'(x_0) < L(x_0) = f(x_0)$, contradicting the convexity of f.
So it must be that $L(x) \leq f(x)\ \forall x < x_0$.

Similarly one proves that $L(x) \leq f(x)\ \forall x > x_2$, so that $px + b - f(x) = L(x) - f(x) \leq 0\ \forall x \notin [x_0, x_2]$.

This gives us

$$\begin{aligned}
f^c(p) &= \sup_x (px - f(x)) \\
&= \sup_x ((px + b) - f(x) - b) \\
&= \sup_x ((px + b) - f(x)) - b \\
&= \sup_{x \in [x_0, x_2]} ((px + b) - f(x)) - b \\
&= M - b \\
&< \infty
\end{aligned}$$

Thus, there is some p with $f^c(p) < \infty$, and this concludes the proof that f^c is a convex function. □

It is also natural to look at the subset of $\mathbb{R}^2$ consisting of all the points above the graph of a given convex function. Moreover, we will need this concept for subsequent proofs.

Definition 5 *Let f be any function. The epigraph of f is the set $[f] \subseteq \mathbb{R} \times \mathbb{R}$ defined by*

$$[f] := \{(x, z) \in \mathbb{R}^2 | z \geq f(x)\}$$

It is straightforward to see that the epigraph of a convex function is a convex set in $\mathbb{R}^2$.

We also need the notion of conjugate sets:

Definition 6 *For any set $M \subseteq \mathbb{R}^2$, the set*

$$M^c := \{(y, w) \in \mathbb{R}^2 | yx - w \leq z \ \forall (x, z) \in \mathbb{R}^2\}$$

is called the <u>*upper conjugate set*</u> *to* M.

There is a concept of lower conjugate set, but we will not need this, so there will be no ambiguity in simply referring to such an M^c as <u>the</u> conjugate set to M.

Note that

$$(y,w) \in M^c \;\Leftrightarrow\; M \subseteq H(y,w) := \{(x,z) | yx - w \leq z\}$$

In other words, asserting that $(y,w) \in M^c$ is the same as saying that the line $z = yx - w$ (the line with slope y and vertical intercept w) lies below or on the boundary of the set M (i.e. every point on this line is either below or on M).

The conjugate sets that will be of interest to us will be those of epigraphs $M = [f]$ of convex functions f.
One would think there would be a nice relation between the conjugate set to an epigraph and the epigraph of a conjugate function. Indeed, there is:

Lemma 3 $[f]^c = [f^c]$

Proof 1) $[f^c] \subseteq [f]^c$:
Take any $(y,w) \in [f^c]$. So $w \geq f^c(y) = \sup_x\{yx - f(x)\} \geq yx - f(x)$ for any x.
We'll show that $[f] \subseteq H(y,w) = \{(x,z) | yx - z \leq w\}$.
Take any $(x,z) \in [f]$. Then $z \geq f(x)$, whence $yx - z \leq yx - f(x) \leq w$, $\Rightarrow (x,z) \in H(y,w)$.

2) $[f]^c \subseteq [f^c]$:

Take $(y, w) \in [f]^c$. So $[f] \subset H(y, w) = \{(x, z) | yx - z \leq w\}$.
We need to show that $(y, w) \in [f^c]$; i.e. that $w \geq f^c(y) = \sup_x\{yx - f(x)\}$.
Take any x. Then $(x, f(x)) \in [f]$, so $yx - f(x) \leq w$ by assumption. But since x was arbitrary, this means that $f^c(y) = \sup_x\{yx - f(x)\} \leq w$. Thus $(y, w) \in [f^c]$. $\square$
To get a geometric interpretation of the conjugate function, all we must do is resort to its definition:

Lemma 4 *If f is any function, and y_0 is any point such that $w_0 = f^c(y_0)$ is finite, then $[f] \subseteq H(y_0, w_0)$ and w_0 is the smallest number w such that $[f] \subseteq H(y_0, w)$.*

Proof For the first part, take any $(x', z) \in [f]$. Then $z \geq f(x')$. Since $w_0 = \sup_x(y_0x - f(x))$, we have $w_0 \geq y_0x' - f(x')$, which is not ambiguous (even if $f(x') = \infty$) because $|w_0| < \infty$. So $y_0x' - w_0 \leq f(x') \leq z$, $\Rightarrow (x', z) \in H(y_0, w_0)$.

For the second part, assume that we had some $w < w_0$ such that $[f] \subseteq H(y_0, w)$. Again since $w_0 = \sup_x(y_0x - f(x))$, there must be some x'' such that $w < y_0x'' - f(x'') \leq w_0$, whence $y_0x'' - w > f(x'')$. So $(x'', f(x'')) \notin H(y_0, w)$, but $(x'', f(x'')) \in [f]$; contradicting the first part of the result. $\square$
Saying that $[f] \subseteq H(y_0, w)$ means that the line $z = y_0x - w$ lies below or on the graph of f (i.e. every point on this line is either below or on the graph of f). Thus, the first part of the above lemma says that the line $z = y_0x - f^c(y_0)$ lies entirely below or on the graph of f.
Decreasing the value of the vertical-intercept w of the line $z = y_0x - w$ has the effect of pushing up the line, so the 2nd part of the lemma says that we cannot push the line up any more and still have it lie entirely below or on the graph of f. In other words, the line must meet the

graph of f at some point.
Moreover, if f is convex, then the line cannot meet the graph of f at more than one point, for if it did, then the line segment joining these two points would lie about the graph of f, contradicting the fact that $[f] \subseteq H(y_0, f^c(y_0))$. If f is also differentiable at that point $(x, f(x))$ where the line and graph meet, then this means that the line is tangent to the graph of f at $(x, f(x))$:

Result The above lemma, in the case where f is convex and differentiable.

Proof Take a point y_0 and suppose that $f^c(y_0)$ is finite. Consider the line $z = y_0 x + b$, and suppose it is tangent to the graph of f at the point $(x_0, f(x_0))$. From our observations above, we need to show that $b = -f^c(y_0)$.
Since the function $y_0 x - f(x)$ is differentiable, the value $f^c(y_0) = \sup_x(y_0 x - f(x)) = \max_x(y_0 x - f(x))$ occurs at a point x_0 where $\frac{d}{dx}(y_0 x - f(x))|_{x=x_0} = 0$ - i.e. where $f'(x_0) = y_0$. Then we have

$$\begin{aligned} f^c(y_0) &= (y_0 x_0 - f(x_0)) \\ \Rightarrow b &= f(x_0) - y_0 x_0 \\ &= -f^c(y_0) \end{aligned}$$

$\square$

Note that the first method used to obtain this geometric view of convex conjugate functions is more general because it does not require differentiability of the function. However the second method is constructive; we will use it to explicitly calculate the function f^c for some functions f later on.

We now begin the steps to proving that the convex conjugate transform is an involution on convex functions.

Definition 7 *Let $a, b \in \mathbb{R}$.*
A halfspace of $\mathbb{R}^2$ is any set of the form

$$\begin{aligned} &\{(x,z) \in R^2 | ax + b \leq z\} \ \textit{(nonvertical halfspace)} \\ \textit{or} \ &\{(x,z) \in R^2 | ax \leq b\} \ \textit{(vertical halfspace).} \end{aligned}$$

Definition 8 *Let $M \subseteq R^2$. M is called* <u>*strongly closed*</u> *if M can be written as an intersection of nonvertical halfspaces.*

Theorem 1 *Any closed, convex epigraph $[f]$ is strongly closed if $[f]^c \neq \emptyset$.*

Proof
$[f]$ is an intersection of halfspaces ([6]):

$$[f] = \cap_{i \in I} H_i$$

some of them possibly vertical. We will show that we can discard those vertical halfspaces.

By assumption, there exists a point $(y, w) \in [f]^c$. As we saw before, this means that $[f] \subseteq H(y, w) = \{(x, z) | yx - w \leq z\}$. So we may replace the vertical halfspaces H_i with sets of the form $H \cap H(y, w)$, where H is a vertical halfspace with $[f] \subseteq H$. Take any such $H = \{(x, z) \in \mathbb{R}^2 | ax \leq b\}$. It suffices to show that $H \cap H(y, w)$ is an intersection of nonvertical halfspaces. The idea is to consider halfspaces bounded by lines of steeper and steeper slope all lying below the graph of f, and then to take the intersection of all of these halfspaces.

Specifically, consider the sets $H_\alpha := \{(x,z)|(y+a\alpha)x - w \leq z + b\alpha\}$, $\alpha \geq 0$. We will show that

$$H \cap H(y,w) = \cap_{\alpha \geq 0} H_\alpha \tag{1}$$

First let $(x,z) \in H \cap H(y,w)$. Then for any $\alpha \geq 0$ we have

$$\begin{aligned}
(y+a\alpha)x - w &= yx - w + a\alpha x \\
&\leq z + a\alpha x \ \ \text{(since } (x,z) \in H(y,w)\text{)} \\
&\leq z + b\alpha \ \ \text{(since } ax \leq b \text{ and } \alpha \geq 0\text{)} \\
\Rightarrow (x,z) \in H_\alpha &
\end{aligned}$$

Since $\alpha \geq 0$ was arbitrary, we have $(x,z) \in \cap_{\alpha \geq 0} H_\alpha$, $\Rightarrow H(y,w) \subseteq \cap_{\alpha \geq 0} H_\alpha$. Therefore $H \cap H(y,w) \subseteq \cap_{\alpha \geq 0} H_\alpha$.

For the reverse inclusion, let $(x,z) \in \cap_{\alpha \geq 0} H_\alpha$. Since $H_0 = H(y,w)$ and $\cap_{\alpha \geq 0} H_\alpha \subseteq H_0$, we have $(x,z) \in H(y,w)$.
It remains to show that $(x,z) \in H$ - i.e. that $ax \leq b$.
Notice that

$$\begin{aligned}
&(y+a\alpha)x - w \leq z + b\alpha \\
\Rightarrow &a\alpha x - b\alpha \leq w + z - yx \\
\Rightarrow &ax - b \leq \frac{1}{\alpha}\underbrace{(w+z-yx)}_{\text{fixed}} \ \ \text{for any } \alpha > 0
\end{aligned}$$

Now if $w + z - yx \leq 0$, then clearly $ax - b \leq 0$.
If $w + z - yx \geq 0$, then since $\frac{1}{\alpha}(w+z-yx) \searrow 0$ as $\alpha \to \infty$, we must have $ax - b \leq 0$.
So in any case, $ax - b \leq 0$, $\Rightarrow (x,z) \in H(y,w)$.

Thus $\cap_{\alpha\geq 0}H_\alpha \subseteq H \cap H(y,w)$, so $[f] = \cap_{i\in I}H_i = \cap_{i\in I}[\cap_{\alpha\geq 0}H_\alpha]$ is an intersection of nonvertical halfspaces. □

Example Take $f(x) = x^2$. Then $[f] = \cap_x H_x$, where $H_x = \{(x,z)|z \geq 2x\}$; i.e. $[f]$ is the intersection of the halfspaces formed by the tangent lines to the graph of f. □

Example Let f be defined by

$$f(x) := \begin{cases} 0, \text{ if} -1 \leq x \leq 1 \\ \infty, \text{ otherwise} \end{cases}$$

as mentioned before. Then

$$\begin{aligned}
[f] &= \{(x,z)| -1 \leq x \leq 1, z \geq 0\} \\
&= \{x \geq -1\} \cap \{x \leq 1\} \cap \{z \geq 0\} \\
&= (\underbrace{\{-x \leq 1\}}_{\text{v.h.}} \cap \{-z \leq 0\}) \cap (\underbrace{\{x \leq 1\}}_{\text{v.h.}} \cap \{-z \leq 0\})
\end{aligned}$$

We can rewrite this as

$$[f] = (\cap_{\alpha\geq 0}\{(x,z)| -\alpha x \leq z+\alpha\}) \cap (\cap_{\alpha\geq 0}\{(x,z)|\alpha x \leq z+\alpha\})$$

confirming the above result.

The next lemma follows from the definition of conjugate sets.

Lemma 5 *For any set $M \subseteq \mathbb{R}^2$, we have*

1) $M \subseteq N \Rightarrow N^c \subseteq M^c$,

2) $M \subseteq M^{cc}$, *and* $M^{cc} = M \Leftrightarrow M$ *is strongly closed,*

3) $M^{ccc} = M^c$; *i.e.* M^c *is strongly closed*

For instance, to prove (1), note that if $M \subseteq N$ and $(y', w') \in N^c = \{(y,w)|yx - w \leq z\ \forall (x,z) \in N\}$, then in particular (y', w') satisfies $y'x - w' \leq z\ \forall (x,z) \in M$. So $(y', w') \in M^c$. □

Definition 9 *A function f is called strongly closed if $[f]$ is a strongly closed set.*

With this definition, the preceding lemma becomes

Lemma 6

$$1)\ f_1 \geq f_2 \Rightarrow f_2^c \geq f_1^c,$$
$$2)\ f \geq f^{cc},\ \text{and } f^{cc} = f \Leftrightarrow f \text{ is strongly closed,}$$
$$3)\ f^{ccc} = f^c \quad \text{i.e. } f^c \text{ is strongly closed.}$$

The main theorem of the chapter now follows:

Theorem 2 *If f is convex, then $f^{cc} = f$.*

Proof by Lemma 2, f^c is also convex, so $[f^c] \neq \emptyset$. But $[f]^c = [f^c]$ by Lemma 3, so $[f]^c \neq \emptyset$. But then $[f]$ is strongly closed by Theorem 1, whence $f^{cc} = f$ by Lemma 6. □

2.3 Examples of Convex Conjugate Functions

(i) $f(x) = ax + b$. Then $f^c(p) = \sup_x (p-a)x + b$.
If $p - a > 0$, then $\lim_{x \to \infty} (p-a)x + b = \infty$;
If $p - a < 0$, then $\lim_{x \to -\infty} (p-a)x + b = \infty$;
either way we have $f^c(p) = \infty$ unless $p = a$, in which case $(p -$

$a)x + b = b$ for all x, whence $f^c(p) = b$. Therefore

$$f^c(p) = \begin{cases} b \text{ if } p = a \\ \infty \text{ if } p \neq a \end{cases}$$

(ii) $f(x) = |x|$.

If $-1 \leq p \leq 1$, then $f^c(p) = \sup_x px - |x| = 0$.

If $|p| > 1$, then $f^c(p) = \sup_x px - |x| = \infty$.

Therefore

$$f^c(p) = \begin{cases} 0 \text{ if } -1 \leq p \leq 1 \\ \infty \text{ if } p \geq 1 \end{cases}$$

(iii) $f(x) = 3x^2 - 5x + 1$. Since f is differentiable, we can easily compute f^c by noting that for any p, $f^c(p) = \sup_x(px - f(x))$ is finite, and thus is attained at the value $x = x(p)$ such that $p - f'(x) = 0$.

So we set $f'(x) = p$ and get

$$\begin{aligned} 6x - 5 &= p \\ \Rightarrow x(p) &= \frac{p+5}{6} \\ \Rightarrow f^c(p) &= px(p) - f(x(p)) \\ &= p(\frac{p+5}{6}) - 3(\frac{p+5}{6})^2 + 5(\frac{p+5}{6}) - 1 \\ &= \frac{1}{12}(p^2 + 10p + 13) \end{aligned}$$

3 The Functional F_w

3.1 Convexity of F_w

Definition 10 *A function $f : \mathbb{R} \to \mathbb{R}$ is called decreasing if*

$$\forall x, y, \;\; x \leq y \Rightarrow f(x) \geq f(y)$$

Definition 11 *Let $w \geq 0$ be a decreasing function on $[0,1]$ and let $0 \leq t \leq 1$. Define the following function:*

$$F_w(t) := \int_t^1 w(\theta)d\theta$$

Theorem 3 *For any two decreasing functions $v, w \geq 0$ that satisfy the normalization condition $\int_0^1 wd\theta = \int_0^1 vd\theta$, we have*

$$F_w(t) \leq F_v(t) \; \forall t \in [0,1]$$

$$\Leftrightarrow$$

$$\int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta \; \forall x \geq 0$$

The first step to proving this theorem is to show that the inequality constraint in the infimum of the definition of F_w can be replaced by an equality constraint. Then we will show an alternate expression for F_w that will prove to be easier to deal with in the development.

Lemma 7 *For any fixed $t \in [0,1]$ we have*

$$\begin{aligned} F_w(t) &= \inf_{|E|=t} \int_0^1 (1 - \chi_E(\theta))w(\theta)d\theta \quad (\#) \\ \textit{and} &= \inf_{|E|\leq t} \int_0^1 (1 - \chi_E(\theta))w(\theta)d\theta \end{aligned}$$

Proof Notice that this infimum does exist (in the sense of being finite), and in fact, is ≥ 0, because: $w \geq 0$, and $1 - \chi_E \geq 0$ for any such E, therefore the integral is ≥ 0 for any such E.

We will first establish (#). Now, because the integral in the definition of F_w is evaluated from t to 1, it seems more natural to first look at the sets $F \subset [0,1] \ni |F| = 1-t$, i.e. $F = E^c$. Then $\chi_F = 1-\chi_E$, and

$$\int_0^1 (1 - \chi_E(\theta))w(\theta)d\theta = \int_0^1 \chi_F(\theta)w(\theta)d\theta$$

Intuitively, fix a $t \in [0,1]$ and look at the values of the expression $\int_0^1 \chi_F(\theta)w(\theta)d\theta = \int_F w(\theta)d\theta$ for various F with $|F| = 1 - t$.

Since w is decreasing, it appears that the smallest of these values occurs at $F = [t, 1]$:
i.e. $\inf_{|F|=1-t} \int_0^1 \chi_F(\theta)w(\theta)d\theta = \int_t^1 w(\theta)d\theta$.

To formally show this, write $F = A \cup B$, where $A = F \cap [0,t]$ and $B = F \cap [t,1]$. Then we have $1 - t = |F| = |A \cup B| = |A| + |B|$ since $A \cap B = \emptyset$.
Notice that w is bounded on $[0,1]$ because it is decreasing on the closed interval $[0,1]$. In particular, w is bounded above on $[t,1]$ by $c = w(t)$. i.e. $c \geq w(y)$ for all $y \in [t,1]$. Moreover, since w is decreasing, we have $w(x) \geq c$ for any $x \in [0,t]$. Since $A \subseteq [0,t]$, we have in particular $w(x) \geq c$ for all $x \in A$. Thus $w(x) \geq c \geq w(y)$ for all $x \in A$ and all $y \in [0,1]$.

¿From there we get

$$\begin{aligned}
\int_0^1 \chi_F(\theta)w(\theta)d\theta &= \int_F w(\theta)d\theta \\
&= \int_A w(\theta)d\theta + \int_B w(\theta)d\theta \\
&\geq c|A| + \int_B w(\theta)d\theta \\
&= c((1-t) - |B|) + \int_B w(\theta)d\theta \\
&\geq \int_{[t,1]/B} w(\theta)d\theta + \int_B w(\theta)d\theta \\
&= \int_t^1 w(\theta)d\theta
\end{aligned}$$

Since this holds for any $F \ni |F| = 1-t$, we can pass to infimums:

$$\int_t^1 w(\theta)d\theta \leq \inf_{|F|=1-t} \int_0^1 \chi_F(\theta)w(\theta)d\theta$$

But $[t,1]$ is one such F, so we get the reverse inequality:

$$\inf_{|F|=1-t} \int_0^1 \chi_F(\theta)w(\theta)d\theta \leq \int_0^1 \chi_{[t,1]}(\theta)w(\theta)d\theta = \int_t^1 w(\theta)d\theta$$

And so we have

$$\int_t^1 w(\theta)d\theta = \inf_{|F|=1-t} \int_0^1 \chi_F(\theta)w(\theta)d\theta$$

Now put things back in terms of the original sets E with $|E| = t$:

$$\begin{aligned}\int_t^1 w(\theta)d\theta &= \inf_{|F|=1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta \\ &= \inf_{|E|=t}\int_0^1 (1-\chi_E(\theta))w(\theta)d\theta\end{aligned}$$

This proves (#).
We will now show equality of the two infimums. Notice that if $|F| \geq 1-t$, then we can write $F = G \cup H$ where $|G| = 1-t$ and $G \cap H = \emptyset$. Then we have

$$\begin{aligned}\int \chi_F w(\theta)d\theta &= \int \chi_G w(\theta)d\theta + \int \chi_H w(\theta)d\theta \\ &\geq \int \chi_G w(\theta)d\theta \\ &\geq \inf_{|F|=1-t}\int \chi_F w(\theta)d\theta\end{aligned}$$

Since F was arbitrary (with $|F| = 1-t$), we have

$$\inf_{|F|\geq 1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta \geq \inf_{|F|=1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta$$

The reverse inequality is clear, since if $|F| = 1-t$, then $|F| \geq 1-t$:

$$\inf_{|F|\geq 1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta \leq \inf_{|F|=1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta$$

So we now have

$$\inf_{|F|\geq 1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta = \inf_{|F|=1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta$$

and therefore

$$\begin{aligned}
\inf_{|E|\le t}\int_0^1 (1-\chi_E(\theta))w(\theta)d\theta &= \inf_{|F|\ge 1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta \\
&= \inf_{|F|=1-t}\int_0^1 \chi_F(\theta)w(\theta)d\theta \\
&= \inf_{|E|=t}\int_0^1 (1-\chi_E(\theta))w(\theta)d\theta
\end{aligned}$$

as desired. □

Thus, the infimums in the lemma are actually minimums, and are achieved at the particular set $E=[0,t]$.

Now, observe that w is assumed only to be continuous, so F is differentiable with $F_w'(t) = d/dt\int_t^1 w(\theta)d\theta = -w(t) \le 0$. If w itself were differentiable, then F would be twice differentiable, so we would have $F_w''(t) = -w'(t) \ge 0$ since w is decreasing.
In other words, F_w is easily seen to be convex in the particular case that w is differentiable.

However F_w is convex, regardless of whether or not w is differentiable:

Lemma 8 *For any decreasing function $w \ge 0$, F_w is a convex function.*

Proof We need to show that for any $s, t, \lambda \in [0,1]$,

$$F_w(\lambda s + (1-\lambda)t) \le \lambda F_w(s) + (1-\lambda)F_w(t)$$

$$\text{i.e. } \int_{\lambda s+(1-\lambda)t}^1 w(\theta)d\theta \le \lambda\int_s^1 w(\theta)d\theta + (1-\lambda)\int_t^1 w(\theta)d\theta$$

Since F_w is continuous, it is sufficient to prove the above inequality for $\lambda = 1/2$ only. (See Hardy [3]).

We will use the following trivial inequality that comes from the fact that $w \geq 0$ and decreasing If w is defined on an interval $[a, d]$ and if $a \leq b \leq d$ and $b - a = d - b$, then

$$\int_b^d w(\theta)d\theta \leq \int_a^b w(\theta)d\theta \quad (\#)$$

Now, we can assume WLOG that $s < t$. We have

$$\begin{aligned}
\int_{\frac{1}{2}(s+t)}^1 w(\theta)d\theta &= \int_{\frac{1}{2}(s+t)}^t w(\theta)d\theta + \int_t^1 w(\theta)d\theta \\
&= \frac{1}{2}\int_t^1 w(\theta)d\theta + \underbrace{\frac{1}{2}\int_t^1 w(\theta)d\theta + \int_{\frac{1}{2}(s+t)}^t w(\theta)d\theta}_{\text{show that this is } \leq \frac{1}{2}\int_s^1 w(\theta)d\theta}
\end{aligned}$$

But indeed,

$$\begin{aligned}
&\frac{1}{2}\int_t^1 w(\theta)d\theta + \int_{\frac{1}{2}(s+t)}^t w(\theta)d\theta \leq \frac{1}{2}\int_s^1 w(\theta)d\theta \\
\Leftrightarrow &\int_{\frac{1}{2}(s+t)}^t w(\theta)d\theta \leq \frac{1}{2}(\int_s^1 w(\theta)d\theta - \int_t^1 w(\theta)d\theta) \\
\Leftrightarrow 2 &\int_{\frac{1}{2}(s+t)}^t w(\theta)d\theta \leq \int_s^t w(\theta)d\theta
\end{aligned}$$

Now to see why this last inequality is true, we can apply (#) to the intervals $[s, \frac{1}{2}(s+t)]$ and $[\frac{1}{2}(s+t), t]$ (which are of equal length $\frac{1}{2}(t-s)$):

$$\begin{aligned} 2\int_{\frac{1}{2}(s+t)}^{t} w(\theta)d\theta &= \int_{\frac{1}{2}(s+t)}^{t} w(\theta)d\theta + \int_{\frac{1}{2}(s+t)}^{t} w(\theta)d\theta \\ &\leq \int_{s}^{\frac{1}{2}(s+t)} w(\theta)d\theta + \int_{\frac{1}{2}(s+t)}^{t} w(\theta)d\theta \quad \text{(by (\#))} \\ &= \int_{s}^{t} w(\theta)d\theta \end{aligned}$$

Putting all this together, we get

$$\begin{aligned} \int_{\frac{1}{2}(s+t)}^{1} w(\theta)d\theta &\leq \frac{1}{2}\int_{s}^{1} w(\theta)d\theta + \frac{1}{2}\int_{t}^{1} w(\theta)d\theta \\ \text{i.e.}\ \ F_w(\frac{1}{2}(s+t)) &\leq \frac{1}{2}F_w(s) + \frac{1}{2}F_w(t), \text{for any } s, t \in [0,1] \end{aligned}$$

Thus, F_w is convex. □

3.2 The T-transform and the Majorization Property of F_w

Theorem 4 *For any two decreasing functions $v, w \geq 0$ that satisfy the normalization condition $\int_0^1 w d\theta = \int_0^1 v d\theta$, we have*

$$F_w(t) \leq F_v(t)\ \forall\, t \in [0,1]$$

$$\Leftrightarrow$$

$$\int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta\ \forall\, x \geq 0$$

We shall present two similar proofs of this result. The motivation for both methods of proof, suggested by Klemes [5], is the convex conjugate transform. One proof will be via the T-transform, followed by the t-transform, and the other will be by first observing that the T-transform is actually a reparametrized (ordinary) convex conjugate transform, and then invoking involutivity of the convex conjugate transform(which we established in Section 1).

Remark: For any $x \geq 0$, $\inf_{0 \leq t \leq 1}(F_w(t) + xt)$ is really a minimum because $F_w(t) + xt$ is a continuous function of t. We will bare this in mind, but continue to write it as an infimum, in keeping with the more general setting that we shall deal with later on.

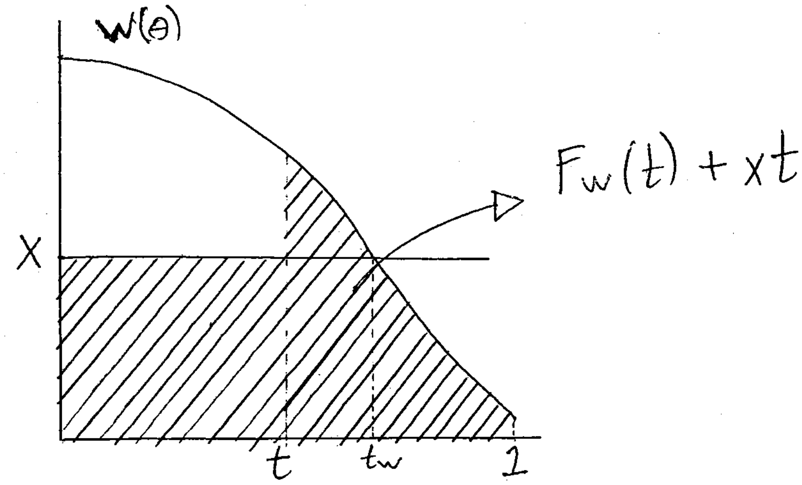
Proof via the T and t-transforms

($\Rightarrow$) Assume that $F_w(t) \leq F_v(t)$ for all $t \in [0,1]$. Consider the following transform of (both sides of) this inequality:

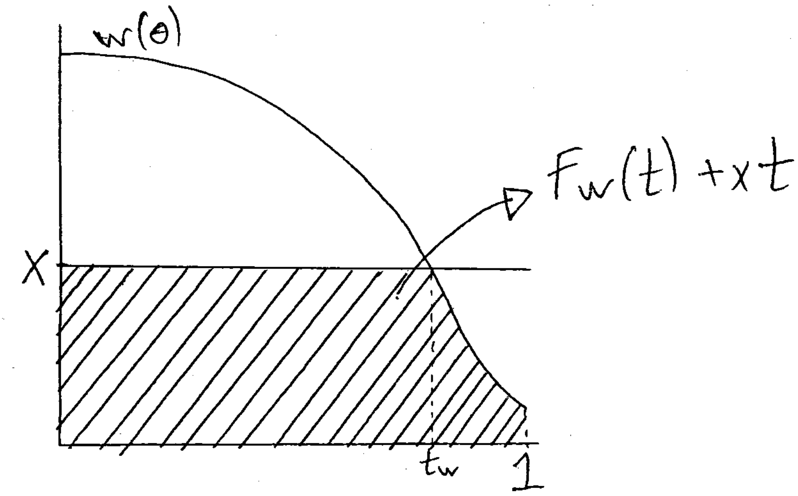
$$\underbrace{\inf_{0 \leq t \leq 1}(F_w(t) + xt)}_{G_w(x)} \leq \underbrace{\inf_{0 \leq t \leq 1}(F_v(t) + xt)}_{G_v(x)}$$

Since $x \geq 0$ was arbitrary, we have $G_w(x) \leq G_v(x)$ for all $x \geq 0$.

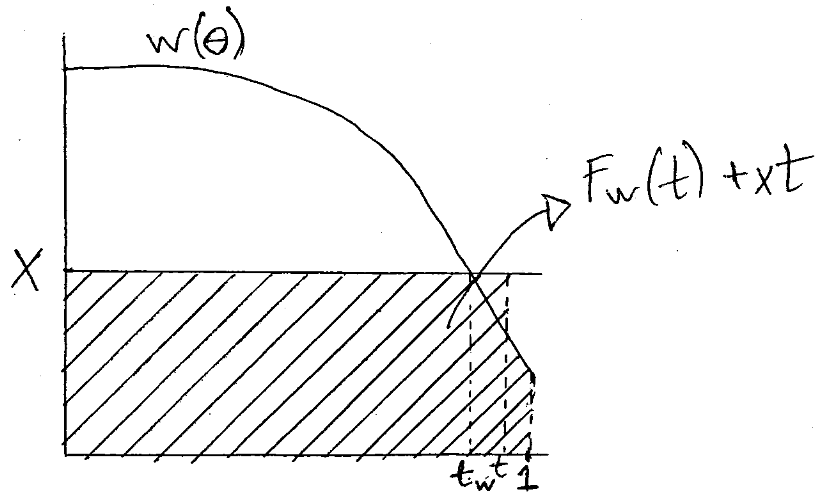
Now, let t_w be the value of t such that $w(t_w) = x$. We will show that for any $x \geq 0$, $\inf_{0 \leq t \leq 1}(F_w(t) + xt)$ is achieved at $t = t_w$.
i.e.: $G_w(x) = \inf_{0 \leq t \leq 1}(F_w(t) + xt) = F_w(t_w) + xt_w$.
This is suggested by the pictures on the following page, which show the quantities $F_w(t) + xt$ for $t \leq t_w$, $t = t_w$, and $t \geq t_w$.



w(θ)
X
Fw(t) + xt
t
tw
1



w(θ)
X
Fw(t) + xt
tw
1



w(θ)
X
Fw(t) + xt
tw
t
1

Formally, we have

1) For $t \geq t_w$:

$$\begin{aligned}
F_w(t) + xt &= \int_t^1 w(\theta)d\theta + xt \\
&= \int_t^1 w(\theta)d\theta + xt_w + x(t - t_w) \\
&= \int_t^1 w(\theta)d\theta + xt_w + \int_{t_w}^t xd\theta \\
&= \int_t^1 w(\theta)d\theta + xt_w + \int_{t_w}^t w(\theta)d\theta + \int_{t_w}^t (x - w(\theta))d\theta \\
&= \int_{t_w}^1 w(\theta)d\theta + xt_w + \int_{t_w}^t (x - w(\theta))d\theta \\
&= F_w(t_w) + xt_w + \int_{t_w}^t (x - w(\theta))d\theta
\end{aligned}$$

But $w(\theta) \leq w(t_w) = x$ for all $\theta \in [t_w, t]$ since w is decreasing, and so $x - w(\theta) \geq 0$ for all $\theta \in [t_w, t]$ hence $\int_{t_w}^t (x - w(\theta))d\theta \geq 0$. And so we have $F_w(t) + xt \geq F_w(t_w) + xt_w$.

2) For $t \leq t_w$:

$$\begin{aligned}
F_w(t) + xt &= \int_t^1 w(\theta)d\theta + xt \\
&= \int_t^{t_w} w(\theta)d\theta + \int_{t_w}^1 w(\theta)d\theta + xt \\
&= \int_t^{t_w} xd\theta + \int_t^{t_w} (w(\theta) - x)d\theta + \int_{t_w}^1 w(\theta)d\theta + xt \\
&= x(t_w - t) + \int_t^{t_w} (w(\theta) - x)d\theta + \int_{t_w}^1 w(\theta)d\theta + xt \\
&= xt_w + F_w(t_w) + \int_t^{t_w} (w(\theta) - x)d\theta
\end{aligned}$$

But $w(\theta) - x \geq 0$ for all $\theta \in [t, t_w]$ since w is decreasing, and so $\int_t^{t_w}(w(\theta) - x)d\theta \geq 0$. So we have $F_w(t) + xt \geq F_w(t_w) + xt_w$ yet again.

Thus for any $x \geq 0$, $G_w(x) \geq F_w(t_w) + xt_w$.

But also $G_w(x) = \inf_{0 \leq t \leq 1}(F_w(t) + xt) \leq F_w(t_w) + xt_w$,
so $G_w(x) = F_w(t_w) + xt_w$ for any $x \geq 0$.

Remark: Note that we now have $G_w(x) = F_w(t_w) + xt_w$, which clearly $= \min(w(\theta), x)$. Thus for any $x \geq 0$, we have $\min(w(\theta), x) \leq \min(v(\theta), x)$ as functions of θ.

Now, back to our objective; We have, for any $x \geq 0$,

$$\begin{aligned}
\int_0^1 (w(\theta) - x)_+ d\theta + G_w(x) &= \int_0^{t_w} (w(\theta) - x) d\theta + F_w(t_w) + x t_w \\
&= \int_0^{t_w} w(\theta) d\theta - x t_w + \int_{t_w}^1 w(\theta) d\theta + x t_w \\
&= \int_0^1 w(\theta) d\theta
\end{aligned}$$

The same is of course true of the function v, so this gives us

$$\begin{aligned}
\int_0^1 (w(\theta) - x)_+ d\theta + G_w(x) &= \int_0^1 w(\theta) d\theta \\
\int_0^1 (v(\theta) - x)_+ d\theta + G_v(x) &= \int_0^1 v(\theta) d\theta
\end{aligned}$$

Finally, using the normalization condition and the fact that $G_w \leq G_v$ that we showed above, we have

$$\begin{aligned}
\int_0^1 (w(\theta) - x)_+ d\theta &= \int_0^1 w(\theta) d\theta - G_w(x) \\
&= \int_0^1 v(\theta) d\theta - G_w(x) \\
&\geq \int_0^1 v(\theta) d\theta - G_v(x) \\
&= \int_0^1 (v(\theta) - x)_+ d\theta
\end{aligned}$$

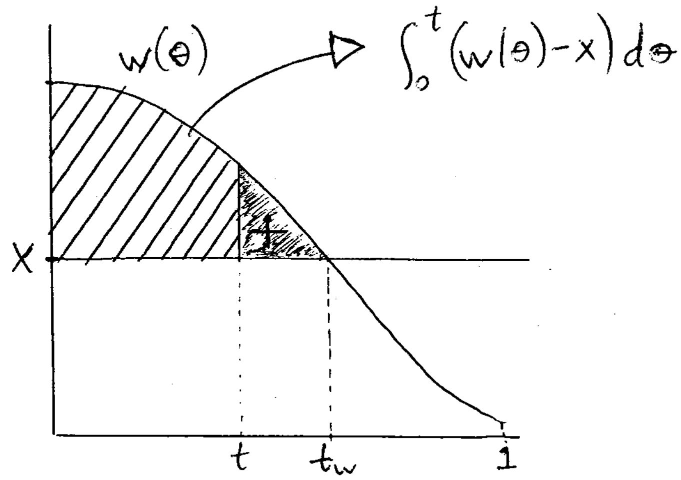
as desired.

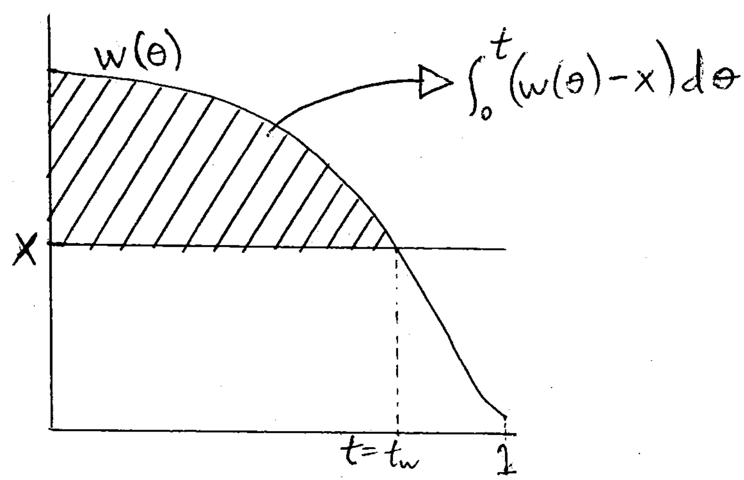
($\Leftarrow$) Assume that $\int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta$ for all $x \geq 0$. Now if we again denote by t_w the point of intersection of the horizontal line x and the function $w(t)$, (i.e.$x = w(t_w)$), we will show that

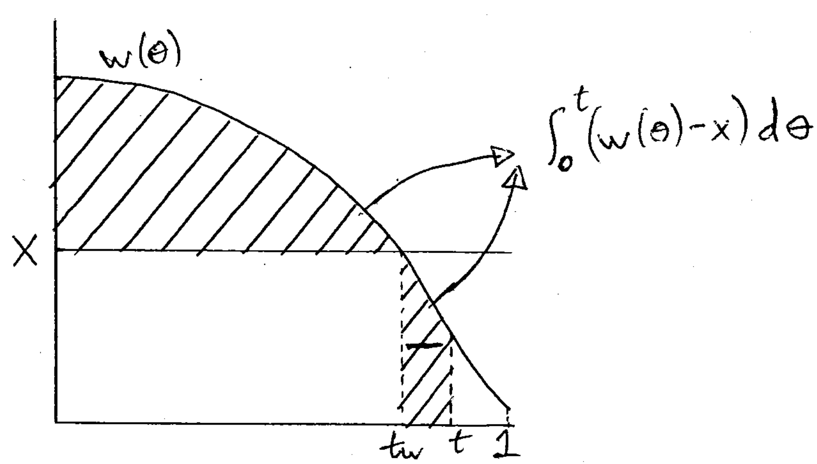
$$\int_0^1 (w(\theta) - x)_+ d\theta, \text{ which } = \int_0^{t_w} (w(\theta) - x) d\theta,$$

$$\text{also } = \sup_{0 \leq t \leq 1} \int_0^t (w(\theta) - x) d\theta$$

Again, this is suggested by the pictures on the following page, which show the quantities $\int_0^t (w(\theta) - x) d\theta$ for $t \leq t_w$, $t = t_w$, and $t \geq t_w$.


$w(\theta)$
$\int_0^t (w(\theta)-x)\,d\theta$
X
+
t
t_w
1


$w(\theta)$
$\int_0^t (w(\theta)-x)\,d\theta$
X
$t = t_w$
1


$w(\theta)$
$\int_0^t (w(\theta)-x)\,d\theta$
X
−
t_w
t
1

Formally, first suppose $t \leq t_w$. Then for all $\theta \in [t, t_w]$ we have $w(\theta) - x \geq 0$ since w is decreasing, and so $\int_t^{t_w}(w(\theta) - x)d\theta \geq 0$. This gives

$$\begin{aligned} \int_0^t (w(\theta) - x)d\theta &\leq \int_0^t (w(\theta) - x)d\theta + \int_t^{t_w} (w(\theta) - x)d\theta \\ &= \int_0^{t_w} (w(\theta) - x)d\theta \end{aligned}$$

Now consider $t \geq t_w$. For all $\theta \in [t_w, t]$ we have $w(\theta) - x \leq 0$ since w is decreasing, and so $\int_{t_w}^t (w(\theta) - x)d\theta \leq 0$. This gives

$$\begin{aligned} \int_0^t (w(\theta) - x)d\theta &= \int_0^{t_w} (w(\theta) - x)d\theta + \int_{t_w}^t (w(\theta) - x)d\theta \\ &\leq \int_0^{t_w} (w(\theta) - x)d\theta \end{aligned}$$

Thus $\int_0^t(w(\theta) - x)d\theta \leq \int_0^{t_w}(w(\theta) - x)d\theta$ for all t,
$\Rightarrow \sup_{0\leq t\leq 1} \int_0^t(w(\theta) - x)d\theta \leq \int_0^{t_w}(w(\theta) - x)d\theta$.

But also $\int_0^{t_w}(w(\theta) - x)d\theta \leq \sup_{0\leq t\leq 1} \int_0^t(w(\theta) - x)d\theta$,
so $\int_0^{t_w}(w(\theta) - x)d\theta = \sup_{0\leq t\leq 1} \int_0^t(w(\theta) - x)d\theta$.

Similarly, of course, for the function v.

Now, let $C = \int_0^1 w(\theta)d\theta$ for ease of notation. For any $x \geq 0$, we have

$$
\begin{aligned}
G_w(x) &= \inf_{0\leq t\leq 1}(F_w(t) + xt) \\
&= \inf_{0\leq t\leq 1}(\int_t^1 w(\theta)d\theta + \int_0^t xd\theta) \\
&= \inf_{0\leq t\leq 1}(\int_0^1 w(\theta)d\theta - \int_0^t w(\theta)d\theta + \int_0^t xd\theta) \\
&= \inf_{0\leq t\leq 1}(C - \int_0^t (w(\theta) - x)d\theta) \\
&= C - \sup_{0\leq t\leq 1} \int_0^t (w(\theta) - x)d\theta
\end{aligned}
$$

This gives

$$
\begin{aligned}
G_w(x) &= C - \sup_{0\leq t\geq 1} \int_0^t (w(\theta) - x)d\theta \\
&= C - \int_0^1 (w(\theta) - x)_+ d\theta \text{ (by the preceding calculation)} \\
&\leq C - \int_0^1 (v(\theta) - x)_+ d\theta \text{ (by assumption)} \\
& C - \sup_{0\leq t\leq 1} \int_0^t (v(\theta) - x)d\theta \text{ (by the preceding calculation)} \\
&= G_v(x)
\end{aligned}
$$

Now, for any fixed $t \in [0, 1]$, consider the following transform of G_w:

$$
\sup_{x\geq 0}\{G_w(x) - xt\}
$$

Intuitively, it appears that this supremum is achieved precisely when $x = x' = w(t)$. But we already showed that for any $x \geq 0$, $G_w(x) = F_w(t_w) + xt_w$, where $w(t_w) = x$.
So we would then have

$\sup_{x\geq 0}\{G_w(x) - xt\} = G_w(x') - x't = (F_w(t) + x't) - x't = F_w(t)$. Thus, if we took the inequality "$G_w(t) \leq G_v(t)$" from above, and performed this sup transform on it, we would get back "$F_w(t) \leq F_v(t)$", concluding the proof of the theorem.

To formally justify this, first observe that we obviously have

$$F_w(t) = G_w(x') - x't \leq \sup_{x\geq 0}\{G_w(x) - xt\}$$

It remains to prove the converse inequality. If $x \leq x'$, i.e. $w(t_w) \leq w(t)$ in the notation we have been using, then $t \leq t_w$ since w is decreasing, and we have

$$\begin{aligned} G_w(x) - xt &= [F_w(t_w) + xt_w] - xt \\ &= \int_{t_w}^{1} w(\theta)d\theta + x(t_w - t) \end{aligned}$$

But $w(\theta) - x \geq 0$ for $\theta \in [t, t_w]$, since $w \searrow$; therefore we get

$$\begin{aligned} G_w(x) - xt &\leq \int_{t_w}^{1} w(\theta)d\theta + x(t_w - t) + \int_{t}^{t_w} [w(\theta) - x]d\theta \\ &= \int_{t_w}^{1} w(\theta)d\theta + x(t_w - t) + \int_{t}^{t_w} w(\theta)d\theta - \int_{t}^{t_w} xd\theta \\ &= \int_{t_w}^{1} w(\theta)d\theta + x(t_w - t) + \int_{t}^{t_w} w(\theta)d\theta - x(t_w - t) \\ &= \int_{t_w}^{1} w(\theta)d\theta + \int_{t}^{t_w} w(\theta)d\theta \\ &= \int_{t}^{1} w(\theta)d\theta \\ &= F_w(t) \end{aligned}$$

If $x' \leq x$, i.e $w(t) \leq w(t_w)$, then $t_w \leq t$ since w is decreasing, and we have

$$\begin{aligned} G_w(x) - xt &= [F_w(t_w) + xt_w] - xt \\ &= \int_{t_w}^{1} w(\theta)d\theta - x(t - t_w) \\ &= \int_{t_w}^{1} w(\theta)d\theta - [\int_{t_w}^{t} w(\theta)d\theta + \int_{t_w}^{t} [x - w(\theta)]d\theta] \\ &= \int_{t_w}^{1} w(\theta)d\theta - \int_{t_w}^{t} w(\theta)d\theta - \int_{t_w}^{t} [x - w(\theta)]d\theta \end{aligned}$$

But $x - w(\theta) \geq 0$ for $\theta \in [t_w, t]$, since w is decreasing, and so we have

$$\begin{aligned} G_w(x) &\leq \int_{t_w}^{1} w(\theta)d\theta - \int_{t_w}^{t} w(\theta)d\theta \\ &= \int_{t}^{1} w(\theta)d\theta \\ &= F_w(t) \end{aligned}$$

so again get $G_w(x) - xt \leq F_w(t)$. Therefore $\sup_{x \geq 0}\{G_w(x) - xt\} \leq F_w(t)$, which is what we needed to show.

Finally, as mentioned up to, we now have, for any $t \in [0, 1]$

$$\begin{aligned} &G_w(t) \leq G_v(t) \text{ from above} \\ \Rightarrow &\sup_{x \geq 0}\{G_w(x) - xt\} \leq \sup_{x \geq 0}\{G_v(x) - xt\} \\ \Rightarrow &F_w(t) \leq F_v(t) \end{aligned}$$

$\square$

Remark As we have just seen, the two main tools that we used in the above proof were two functions that are very similar to the convex conjugate transform; we will now formally define them.

Definition 12 *For any function $f : [0,1] \to \mathbb{R}$, the T-transform of f is the function G defined by*

$$G(p) := \inf_{0 \leq x \leq 1} \{f(x) + px\}, \quad (p \geq 0)$$

For any function $H : \mathbb{R} \to \mathbb{R}$, the t-transform of H is the function F defined by:

$$F(t) := \sup_{p \geq 0} \{H(p) - pt\}, \quad (t \in [0,1])$$

Remark $F \circ G$ is well-defined because $Range(G) = \mathbb{R} = Dom(F)$. Later we will show that $F \circ G = Id$ on any function $f : [0,1] \to \mathbb{R}$, and we will use this to prove the main results of the paper.

Remark One might wonder if the function $G(x)$ really is just an ordinary convex conjugate transform of $F(t)$. Indeed, if we denote by $L(F_w(p))$ the convex conjugate transform of F_w at p, then we have

$$\begin{aligned} G_w(x) &= \inf_{0 \leq t \leq 1} F_w(t) + xt \\ &= -\sup_{0 \leq t \leq 1} -F_w(t) - xt \\ &= -\sup_{0 \leq t \leq 1} (-x)t - F_w(t) \\ &= -L(F_w(-x)) \end{aligned}$$

Now, we can extend $G_w(x)$ to all of $\mathbb{R}$ (which is how Stoer defines the convex conjugate transform [7]) by extending F_w linearly with slope $p = -x$; moreover, we can just as easily make F_w continuous on $\mathbb{R}$:

Let $F_w(t) := F_w(0) + pt$ for $t < 0$,
and $F_w(t) := pt + (F_w(1) - p)$ for $t > 1$ (so that F_w goes through the point $(1, F_w(1))$. See picture next page.
This gives us

$$\text{for } t < 0\text{: } pt - F_w(t) = p(0) - F_w(0) \leq \sup_{0 \leq t \leq 1} pt - F_w(t)$$
$$\text{for } t > 1\text{: } pt - F_w(t) = p(1) - F_w(1) \leq \sup_{0 \leq t \leq 1} pt - F_w(t)$$

Therefore

$$pt - F_w(t) \leq \sup_{0 \leq t \leq 1} pt - F_w(t) \text{ for all } t \in \mathbb{R}$$
$$\Rightarrow \sup_{t \in \mathbb{R}} pt - F_w(t) \leq \sup_{0 \leq t \leq 1} pt - F_w(t)$$

And clearly we have $\sup_{t \in \mathbb{R}} pt - F_w(t) \geq \sup_{0 \leq t \leq 1} pt - F_w(t)$, threrefore $\sup_{t \in \mathbb{R}} pt - F_w(t) = \sup_{0 \leq t \leq 1} pt - F_w(t)$,
which extends G_w to a sup over all $t \in \mathbb{R}$.

Therefore we really can look at G_w as a convex conjugate transform of F_w, and we can prove the above theorem this way, by invoking involutivity of the convex conjugate transform that we established in the last chapter.

Proof of Theorem 4 using the convex conjugate transform:
($\Rightarrow$) Fix any $p \leq 0$. Then

$$\begin{aligned} &F_w(t) \leq F_v(t) \, \forall t \in [0,1] \\ \Rightarrow &F_w(t) \leq F_v(t) \, \forall t \in \mathbb{R} \end{aligned}$$

since our above extension of F_w,F_v to all of $\mathbb{R}$ preserves $\leq$ in this relation. But then we have

$$\begin{aligned} &\Rightarrow pt - F_w(t) \geq pt - F_v(t) \forall t \in \mathbb{R} \\ &\Rightarrow \sup_{t \in \mathbb{R}}[pt - F_w(t)] \geq \sup_{t \in \mathbb{R}}[pt - F_v(t)] \\ &\Rightarrow \sup_{t \in [0,1]}[pt - F_w(t)] \geq \sup_{t \in [0,1]}[pt - F_v(t)] \text{ (by extension of } G \text{ above)} \\ &\Rightarrow -\sup_{t \in [0,1]}[pt - F_w(t)] \leq -\sup_{t \in [0,1]}[pt - F_v(t)] \\ &\Rightarrow G_w(-p) \leq G_v(-p) \text{ (we showed this in the remark above)} \\ &\Rightarrow \int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta \;\; (x = -p \geq 0) \end{aligned}$$

where the last implication was shown in the first part of the above proof of theorem 3.
Since $x = -p \geq 0$ was arbitrary, we see that
$\int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta \; \forall x \geq 0$.

($\Leftarrow$) Fix any $x \geq 0$. Then

$$\int_0^1 (w(\theta) - x)_+ d\theta \geq \int_0^1 (v(\theta) - x)_+ d\theta$$

$$\Rightarrow G_w(x) \leq G_v(x)$$

(shown in the first part of the above proof of theorem 3)

$$i.e. \; -L(F_w(-x)) \leq -L(F_v(-x)) \text{ (we showed this above)}$$

$$\Rightarrow L(F_w(p)) \geq L(F_v(p)) \text{ (where } p = -x\text{)}$$

$$\Rightarrow tp - L(F_w(p)) \leq tp - L(F_v(p)) \text{ (for any } t \in \mathbb{R}\text{)}$$

$$\Rightarrow \sup_{t \in \mathbb{R}}[tp - L(F_w(p))] \leq \sup_{t \in \mathbb{R}}[tp - L(F_v(p))]$$

$$i.e. \; L[L(F_w(p))(t)] \leq L[L(F_v(p))(t)]$$

But F_w, F_v are convex by Theorem (8) so, by Theorem (2) L is an involution on each of them. Therefore, the preceding line becomes $F_w(t) \leq F_v(t)$. Moreover, since $t \in \mathbb{R}$ was arbitrary, this completes the proof. $\square$

4 The Functional J_w and the Majorization Property of J_w

It is natural to consider the slightly more general functional

$$J_w : [0.1] \to \mathbb{R},$$
$$J_w(t) := \inf_{\|P\|_1 \le t} \int_0^1 |1 - P(\theta)|w(\theta)d\theta$$

where $P \in L^1[0,1]$, and where $w \ge 0$ is bounded but not necessarily decreasing.
To compare J_w to F_w (which required w to be decreasing), we can consider the decreasing rearrangement P^* of P and w^* of w, whence

$$\int_0^1 |1 - P(\theta)|w(\theta)d\theta = \int_0^1 |1 - P^*(\theta)|w^*(\theta)d\theta$$

The preceding equality of integrals is left as an exercise in the theory of decreasing rearrangements of functions (see [2]). We readily see that if P is a characteristic function, then $J_w(t) = F_w(t)$ for all $t \in [0,1]$.
The majorization result can be shown for J_w:

$$J_w(t) \le J_v(t) \;\; \forall\, t \in [0,1] \Leftrightarrow \int_0^1 \frac{1}{w(\theta) + \lambda} d\theta \ge \int_0^1 \frac{1}{v(\theta) + \lambda} d\theta \;\; \forall\, \lambda \ge 0$$

However, it is the functional K_w (where P is an L^2 function rather than an L^1 function) that we wish to develop, along the lines of the treatment of F_w last section. Moreover, the results for J_w, including the above majorization property, follow from those for K_w with appropriate minor modifications. Therefore, we omit the proofs of the results for J_w, and develop the treatment in detail with the functional K_w in the next section.

5 The Functional K_w and the Majorization Property of K_w

5.1 Convexity of K_w

Definition 13 *Fix a bounded (not necessarily decreasing) function $w \geq 0$ on $[0,1]$, and define a function $K_w : [0,1] \to \mathbb{R}$ by:*

$$K_w(t) := \inf_{||P||_2^2 \leq t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$$

where P ranges over the real L^2 functions, and $||P||_2$ is the usual L^2 norm of P.

The main result of this section is the following result, discussed by Klemes [5]:

Theorem 5

$$K_w(t) \leq K_v(t) \;\; \forall\, t \in [0,1] \Leftrightarrow \int_0^1 \frac{1}{w(\theta)+\lambda} d\theta \geq \int_0^1 \frac{1}{v(\theta)+\lambda} d\theta \;\; \forall\, \lambda \geq 0$$

To begin to prove this theorem, we'll show, as we did with the function F_w, that we can replace the inequality with equality, in the infimum constraint of K_w. And, again as with F_w, this will be very useful in explicitly calculating the T-transform of K_w, and ultimately proving the above theorem, for K_w as we did for F_w.

Remark: The reason K_w was defined as an infimum with an inequality constraint (instead of an equality constraint) in [5] is because this definition clearly allows us to see that K_w is a decreasing function.

Lemma 9 *For any $t \in [0,1]$, we have*

$$K_w(t) := \inf_{\|P\|_2 \le t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta = \inf_{\|P\|_2 = t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$$

Proof

1) For any fixed $t \in [0,1]$, there are more analytic functions P with $\|P\|_2 \le t$ than $\|P\| = t$, so we trivially have

$$\inf_{\|P\|_2 \le t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta \le \inf_{\|P\|_2 = t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$$

2) It remains to prove that

$$\inf_{\|P\|_2 \le t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta \ge \inf_{\|P\|_2 = t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta \quad (*)$$

For this, it is enough to show that for any P_1 with $\|P_1\|_2 < t$, we can find a P_2 with $\|P_2\| = t$ such that
$\int_0^1 |1 - P_2(\theta)|^2 w(\theta) d\theta \le \int_0^1 |1 - P_1(\theta)|^2 w(\theta) d\theta \quad (\#)$;

for then, if $(*)$ were not true;
i.e. if we had $\inf_{\|P\|_2 \le t} \int_0^1 |1-P(\theta)|^2 w(\theta) d\theta < \inf_{\|P\|_2 = t} \int_0^1 |1-P(\theta)|^2 w(\theta) d\theta$, then we could find a P' with $\|P'\|_2 < t$ such that

$$\int_0^1 |1 - P'(\theta)|^2 w(\theta) d\theta < \inf_{\|P\|_2 = t} \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$$

i.e. $\displaystyle\int_0^1 |1 - P'(\theta)|^2 w(\theta) d\theta < \int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$ for any $P \ni \|P\|_2 = t$,

contradicting $(\#)$.

So let's take any P_1 with $||P_1|| < t$.

First let's assume that $0 \leq P_1 \leq 1$.
Since the norm $||\cdot||_2 : L^2[0,1] \to \mathbb{R}^+$ is a continuous function, the Intermediate Value Theorem tells us that we can find a function $P_1' \in L^2[0,1]$ such that $||P_1'|| = t$. Since $||P_1|| < t < 1$ and $||1|| = 1$, we must have $P_1 < P_1' \leq 1$. In other words, we can increase P_1 pointwise up to a function P_1' with norm=t, still keeping $P_1' \leq 1$.

Therefore

$$
\begin{aligned}
&|1 - P_1'| = 1 - P_1' \leq 1 - P_1 = |1 - P_1|, \\
\Rightarrow &\int_0^1 |1 - P_1'(\theta)|^2 w(\theta) d\theta \leq \int_0^1 |1 - P_1(\theta)|^2 w(\theta) d\theta
\end{aligned}
$$

Since P_1 was arbitrary (with norm $< t$), (#) is proven.

Now we remove the restriction that $P = P_1 \in [0,1]$.
The idea is doing this is that if P takes on values outside $[0,1]$, then we can rescale P pointwise down to a function in $[0,1]$; in other words, for any $t \in [0,1]$, if $P(t) \leq 1$, then leave $P(t)$ alone, and if $P(t) > 1$, redefine P by $P(t) := 1$. redefining P only decreases $||P||$, and so preserves $||P|| < t$, and more importantly, it also decreases the value of $|1 - P|$, and therefore the value of $\int_0^1 |1 - P(\theta)|^2 w(\theta) d\theta$. So the desired inequality follows from the one just proved for $P \in [0,1]$.
(See picture next page). $\square$

Lemma 10 *The above result also generalizes to the case where P is complex-valued.*

The first inequality $\leq$ in 1) above is just a fact about infimum's, and so is obviously true even for complex-valued P. It remains to show 2).

We can use the inequality $|1-|P|| \leq |1-P|$ (which is good for any P), and then apply 2) to the inf that ranges over the (real-valued) functions $|P|$:

$$\inf_{||P||_2 \leq t} \int_0^1 |1-P(\theta)|^2 w(\theta)d\theta \geq \inf_{||P||_2 \leq t} \int_0^1 |1-|P(\theta)||^2 w(\theta)d\theta$$
$$= \inf_{|||P|||_2 = ||P|| = t} \int_0^1 |1-|P(\theta)||^2 w(\theta)d\theta$$
(by Lemma 9, since $|P| \in \mathbb{R}$)

$$\geq \inf_{||P||_2 = t} \int_0^1 |1-P(\theta)|^2 w(\theta)d\theta,$$

where this last line follows from the fact that the set
$\{P : \mathbb{C} \to \mathbb{R}\} \subseteq \{P : \mathbb{C} \to \mathbb{C}\}$.
This completes the proof of the lemma. □
Once we establish that K_w is convex, we can apply the T-transform to it, and then the t-transform to invert it.
We give two proofs of convexity of K_w; one relies on using the infimum of the definition of K_w itself, and two norm inequalities (one for the inf constraint and one for K_w); and the other proof first computes this infimum.

Theorem 6 *$H = K_w$ is a convex function.*

Proof I
This is the more direct proof. We need to show that $H(pt_1+(1-p)t_2) \leq pH(t_1) + (1-p)H(t_2)$ for any $p, t_1, t_2 \in [0, 1]$.

We can assume that $p \neq 0, 1$, since we trivially have equality in either of those cases.

Now for any t, the integral in the definition of H is a continuous function of g, so the inf is actually attained by some function g_1 (i.e. the inf is really a min):

i.e. for some function g_1 (with, necessarily $||g_1|| = t_1$ by Lemma 3), we have

$$H(t_1) = \inf_{||g||_2^2 = t_1} \int_0^1 (1-g)^2 w = \int_0^1 (1-g_1) w$$

Likewise for $H(t_2)$: for some function g_2 with $||g_2|| = t_2$, we have

$$H(t_1) = \inf_{||g||_2^2 = t_1} \int_0^1 (1-g)^2 w = \int_0^1 (1-g_1) w$$

Let $t_3 = pt_1 + (1-p)t_2$. We only need ONE function $g_3 \in L_2$ with $||g_3||_2^2 = t_3$ and such that $\int_0^1 (1-g_3) w \leq p \int_0^1 (1-g_1)^2 w + (1-p) \int_0^1 (1-g_2)^2 w$, for then we'd have

$$\begin{aligned}
H(t_3) &= \inf_{||g||_2^2 \leq t_3} \int_0^1 (1-g)^2 w \\
&\leq \int_0^1 (1-g_3)^2 w \\
&\leq p \int_0^1 (1-g_1)^2 w + (1-p) \int_0^1 (1-g_2)^2 w \\
&= pH(t_1) + (1-p) H(t_2)
\end{aligned}$$

as desired

Maybe a simple choice like $g_3 = pg_1 + (1-p)g_2$ would work. Is this g_3 even a candidate, though?

Clearly $g_3 \in L^2$ since $g_1, g_2 \in L^2$ and L^2 is a vector space. To show that

$\|g_3\|_2^2 \leq t_3$, observe that $(pg_1 + (1-p)g_2)^2 \leq pg_1^2 + (1-p)g_2^2$ because the function $f(x) = x^2 : x \in \mathbb{R}$ is a convex function.
So we have

$$\begin{aligned}
\|g_3\|_2^2 &= \int_0^1 g_3^2 \\
&= \int_0^1 (pg_1 + (1-p)g_2)^2 \\
&\leq \int_0^1 pg_1^2 + (1-p)g_2^2 \\
&= p\int_0^1 g_1^2 + (1-p)\int_0^1 g_2^2 \\
&= p\|g_1\|_2^2 + (1-p)\|g_2\|_2^2 \\
&= pt_1 + (1-p)t_2 \\
&= t_3
\end{aligned}$$

Now we need to show that

$$\int_0^1 (1 - \underbrace{(pg_1 + (1-p)g_2)}_{g_3})^2 w \leq p\int_0^1 (1-g_1)^2 w + (1-p)\int_0^1 (1-g_2)^2 w$$

But indeed we have $(1-(pg_1+(1-p)g_2))^2 \leq p(1-g_1)^2+(1-p)(1-g_2)^2$ by convexity of the function $f(x) = (1-x)^2 : x \in \mathbb{R}$;
We then multiply by $w \geq 0$ and take integrals of both sides of this inequality, thus obtaining the above inequality.

Therefore, we have $H(t_3) \leq pH(t_1)+(1-p)H(t_2)$ for any $t_1, t_2, p \in [0,1]$, proving that H is a convex function. $\square$

Proof II

The main idea of this proof is the triangle inequality from two norms:

one is the $L^2([0,1],\mathcal{M},\text{Lebesgue})$ norm with the infimum constraint in the definition of H: $\|g+h\|_2 \leq \|g\|_2 + \|h\|_2$;

For the other norm, observe that for any $f \in L^2$, we have $f^2 \in L^1$, and since w is bounded, we have $f^2 w \in L^1$. Therefore we can define

$$\|\cdot\|_w : L^2 \to [0,\infty)$$
$$\text{by } \|f\|_w := \sqrt{\int_0^1 f^2 w}$$

Notice that $\|\cdot\|_w$ is just a special case of the $L^2([0,1],\mathcal{M},\mu)$-norm $\|f\|_2 := \sqrt{\int_0^1 |f|^2 d\mu}$, where $d\mu = w d\theta$ and $d\theta$ represents integration with respect to Lebesgue measure.

Now, in order to use this norm, we will have to work not directly with H, but with $\sqrt{H}$, and use the fact that

$$\sqrt{\inf_{\|g\|\leq t} \int_0^1 (1-g)^2 w} = \inf_{\|g\|\leq t} \underbrace{\sqrt{\int_0^1 (1-g)^2 w}}_{\|1-g\|_w} \tag{2}$$

(1) is easily established, in the more general context that

$$\sqrt{\inf_t f(t)} = \inf_t \sqrt{f(t)}$$

for any function $f \geq 0$.

Indeed, if $t = \inf_t f(t)$, then

$$\begin{aligned} t &\leq f(t) \text{ for all } t \\ \Rightarrow \sqrt{t} &\leq \sqrt{f(t)} \text{ for all } t \\ \Rightarrow \sqrt{\inf_t f(t)} = \sqrt{t} &\leq \inf_t \sqrt{f(t)} \end{aligned}$$

Conversely, if $t = \inf_t \sqrt{f(t)}$ then

$$\begin{aligned} t &\leq \sqrt{f(t)} \text{ for all } t \\ t^2 &\leq f(t) \text{ for all } t \\ \Rightarrow t^2 &\leq \inf_t f(t) \\ \Rightarrow \inf_t \sqrt{f(t)} = t &\leq \sqrt{\inf_t f(t)} \end{aligned}$$

Now, as we did in Part I, we let $g_3 = pg_1 + (1-p)g_2$, where $H(t_1)$ is minimized at g_1 and $H(t_2)$ is minimized at g_2. Let $t_3 = pt_1 + (1-p)t_2$, but this time we will obtain the required norm constraint on g_3 by using the Δ inequality for the $\|\cdot\|_2$ norm:

$$\begin{aligned} \|g_3\|_2 &= \|pg_1 + (1-p)g_2\|_2 \\ &\leq p\|g_1\|_2 + (1-p)\|g_2\|_2 \;\; (\Delta \text{ ineq. of } \|\cdot\|_2) \\ &= p\sqrt{t_1} + (1-p)\sqrt{t_2} \\ &\leq \sqrt{pt_1 + (1-p)t_2} \\ &\text{by concavity of the function } f(x) = \sqrt{x} : x \in [0, \infty) \\ &= \sqrt{t_3}; \\ &\text{i.e. } \|g_3\|_2^2 \leq t_3 \end{aligned}$$

Ultimately, we must show that

$$H(t_3) \leq pH(t_1) + (1-p)H(t_2) \quad (\#)$$

So far, we have

$$\begin{aligned}
\sqrt{H(t_3)} &= \sqrt{\inf_{\|g\|=t_3} \int_0^1 (1-g)^2 w} \\
&\leq \inf_{\|g\|\leq t} \sqrt{\int_0^1 (1-g)^2 w} \\
&\leq \sqrt{\int_0^1 (1-g_3)^2 w} \\
&= \|1-g_3\|_w \\
&= \|1-pg_1-(1-p)g_2\|_w \\
&= \|p(1-g_1)+(1-p)(1-g_2\|_w \\
&\leq p\|(1-g_1)\|_w + (1-p)\|(1-g_2)\|_w \ \ (\Delta \text{ ineq. of } \|\cdot\|_w) \\
&= p\sqrt{\int_0^1 (1-g_1)^2 w} + (1-p)\sqrt{\int_0^1 (1-g_2)^2 w} \\
&= p\sqrt{H(t_1)} + (1-p)\sqrt{H(t_2)}
\end{aligned}$$

Now square both sides. To obtain (#) we need to show that
$[p\sqrt{H(t_1)} + (1-p)\sqrt{H(t_2)}]^2 \leq pH(t_1) + (1-p)H(t_2)$.

But

$$
\begin{aligned}
&[p\sqrt{H(t_1)} + (1-p)\sqrt{H(t_2)}]^2 \leq pH(t_1) + (1-p)H(t_2) \\
&\Leftrightarrow p^2H(t_1) + (1-p)^2H(t_2) + 2p(1-p)\sqrt{H(t_1)}\sqrt{H(t_2)} \\
&\leq pH(t_1) + (1-p)H(t_2) \\
&\Leftrightarrow 2p(1-p)\sqrt{H(t_1)}\sqrt{H(t_2)} \\
&\leq (p-p^2)H(t_1) + [(1-p)-(1-p)^2]H(t_2) \\
&\Leftrightarrow 2p(1-p)\sqrt{H(t_1)}\sqrt{H(t_2)} \leq p(1-p)[H(t_1) + H(t_2)] \\
&\Leftrightarrow 2\sqrt{H(t_1)}\sqrt{H(t_2)} \leq H(t_1) + H(t_2) \text{ since } p \neq 0, 1
\end{aligned}
$$

But for any $a, b \geq 0$, we have

$$
\begin{aligned}
2\sqrt{a}\sqrt{b} \leq a + b \quad &\Leftrightarrow \quad a - 2\sqrt{a}\sqrt{b} + b \geq 0 \\
&\Leftrightarrow (\sqrt{a} - \sqrt{b})^2 \geq 0, \quad \text{which is obviously true.}
\end{aligned}
$$

Combining these facts, we have $H(t_3) \leq [p\sqrt{H(t_1)} + (1-p)\sqrt{H(t_2)}]^2 \leq pH(t_1) + (1-p)H(t_2)$, as desired. This completes the second proof of convexity of H. □

5.2 The T-Transform, and a simplification of the problem of computing it for K_w

Our strategy for using the T-transform to prove Theorem 5: we will be applying it to the function K_w to prove one side of the equivalence, and then apply the t-transform to get back K_w and hence prove the other implication in the equivalence.

Recall the T-transform and the t-transform:

$$\text{T-transform: } G(p) := \inf_{0 \le x \le 1} f(x) + px \ \ (p \ge 0)$$
$$t\text{-transform: } F(t) := \sup_{p \ge 0} G(p) - pt \ \ (t \in [0,1])$$

At first glance, it would seem that we would need to prove that the $F \circ G$ is the identity on any function f, in order to be able to use these transforms to prove the majorization equivalence.
But it suffices to show that $F \circ G$ is the identity on an arbitrary piecewise linear function with negative slopes.

Lemma 11 *Fix $p \ge 0$, and take a point $T \in [0,1]$ where $K_w(x) + px$ attains its minimum (recall that the infimum in $K_w(x) + px$ is a minimum since this a continuous function of x.). In other words, take T such that $\inf_x(K_w(x) + px) = K_w(T) + pT$. Let l_w be a piecewise linear function lying above K_w but meeting K_w at T - i.e. such that $l_w(t) \ge K_w(t)$ for all T and $l_w(T) = K_w(T)$.*

Then the function $x \to l_w(x) + px$ also attains its minimum at $x = T$, where its value agrees with that of $K_w(x)+px$ by the way we constructed l_w.

Proof: First, we have

$$\begin{aligned} &K_w(x) \le l_w(x) \ \forall x \in [0,1] \\ &\quad \Rightarrow K_w(x) + px \le l_w(x) + px \ \forall x \in [0,1] \\ &\quad \Rightarrow \inf_x (K_w(x) + px) \le \inf_x (l_w(x) + px) \end{aligned}$$

But we also have

$$\begin{aligned}\inf_x (l_w(x) + px) &\leq l_w(T) + pT \\ &= K_w(T) + pT \\ &= \inf_x (K_w(x) + px)\end{aligned}$$

Therefore $\inf_x(K_w(x) + px) = \inf_x(l_w(x) + px)$, and since $p \geq 0$ was arbitrary, we see that these two infimums are equal as functions of $p \geq 0$. □

Remark: In practice, there are many ways we could choose line segments to do the job. A natural way is to take the line segments connecting the points $K_w(0)$, $K_w(T)$, and $K_w(T)$,$K_w(1)$, respectively, where T is any point $\in [0, 1]$. The proof that the t-transform inverts the T-transform of this piecewise linear function relies on the essential fact that it lies above the graph of K_w because K_w is convex.

Theorem 7 *Let f be any piecewise linear function with negative slopes. Then the t-transform inverts the T-transform of f. In other words, $(F \circ G)(f) = f$.*

Proof WLOG we may assume that f is linear with negative slope: let $f(x) = a - bx, b \geq 0$. For fixed $p \geq 0$, consider the T-transform of f,

$$\begin{aligned}G(p) &= \inf_{0 \leq x \leq 1} (a - bx) + px \\ &= \inf_{0 \leq x \leq 1} a + (p - b)x\end{aligned}$$

So we look at $a + (p - b)x$ as a function of $x \in [0, 1]$; this is a line segment; and we vary the value of $p \geq 0$. If $p \geq b$, this line has positive slope, thus its minimum is attained at the left endpoint $x = 0$, with

value a.
If $p \leq b$, then $a+(p-b)x$ is a line segment with negative slope, whence the minimum is attained at the right endpoint $x=1$, with value $a+p-b$.

Therefore we have

$$G(p) = \begin{cases} a & \text{if } p \geq b \\ a-b+p & \text{if } p \leq b \end{cases}$$

We now take

$$\begin{aligned} F(t) &= \sup_{p\geq 0} G(p) - pt \\ &= \begin{cases} \sup_{p\geq 0}(a-pt) & \text{if } p \geq b \\ \sup_{p\geq 0}(a-b+p-pt) = \sup_{p\geq 0}(a-b+p(1-t)) & \text{if } p \leq b \end{cases} \end{aligned}$$

So now we look at the piecewise linear function (of $p \geq 0$), $G(p)-pt$, which is $a-pt$ for $p \geq b$, and $(a-b)+p(1-t)$ for $p \leq b$ and we vary the value of $t \in [0,1]$. See picture on next page. Note: In the graphs, $a \geq b$ WLOG.

The first piece ($p \leq b$), $a-b+p(1-t)$ has positive slope for any t, while the second piece ($p \geq b$), $a-pt$, has negative slope for any t. Moreover, it is easily seen that the function $G(p)-pt$ is continuous in p: the only question is continuity at the point b, and here we have

$$\lim_{p\to b^-}(a-b)+p(1-t) = a-b+b(1-t) = a-bt = \lim_{p\to b^+} a-pt$$

Therefore, the maximum value of $G(p) - pt$ is attained at $p = b$, whence

$$\begin{aligned} F(t) &= \sup_{p \geq 0} (G(p) - pt) \\ &= G(b) - bt \\ &= a - bt \quad \text{since } G(b) = a \text{ from above.} \end{aligned}$$

Thus, F is the original function (line) f, which is what we wanted to show. □

Combining the previous lemma and theorem, we see that the t-transform is an inverse of the T-transform, for any convex function.

5.3 The Majorization Property of K_w

Proof of Theorem 5 (page 40)

($\Rightarrow$) Recall that

$$K_w(t) = \inf_{\|g\|_2^2 = t} \int_0^1 (1-g)^2 w(\theta) d\theta \text{ by Lemma 3}$$

Let v, w be functions as in the hypotheses of Theorem 5.

The first step is purely mechanical: assuming that $K_w(t) \leq K_v(t)$ for any $t \in [0,1]$, we take an T-transform of both sides. Fix any $\lambda > 0$.

We have

$$\inf_{0\le t\le 1}(K_w(t)+\lambda t) \le \inf_{0\le t\le 1}(K_v(t)+\lambda t)$$

$$\text{i.e. } \inf_{0\le t\le 1}([\inf_{\|g\|_2^2=t}(1-g)^2 w(\theta)d\theta]+\lambda t) \le \inf_{0\le t\le 1}([\inf_{\|g\|_2^2=t}(1-g)^2 v(\theta)d\theta]+\lambda t)$$

Now, it would be nice to have a more tangible expression for either side on the last line above; perhaps we can consolidate the two infimums into a single infimum, and then actually evaluate it.

For any function g in the inner infimum, we have $t = \|g\|_2^2 = \int_0^1 g(\theta)^2 d\theta$, where $t \in [0,1]$.

Therefore

$$\begin{aligned}
&\inf_{0\le t\le 1}([\inf_{\|g\|_2^2=t}\int_0^1 (1-g(\theta))^2 w(\theta)d\theta]+\lambda t)\\
&=\inf_{0\le t\le 1}([\inf_{\|g\|_2^2=t}\int_0^1 (1-g(\theta))^2 w(\theta)d\theta]+\lambda \int_0^1 g(\theta)^2 d\theta)\\
&=\inf_{0\le t\le 1}(\inf_{\|g\|_2^2=t}\int_0^1 ((1-g(\theta))^2 w(\theta)+\lambda g(\theta)^2)d\theta)\\
&=\inf_{0\le \|g\|_2^2\le 1}\int_0^1 (1-g(\theta))^2 w(\theta)+\lambda g(\theta)^2)d\theta
\end{aligned}$$

since t no longer appears explicitly in the inner infimum

So our above inequality involving v, w becomes

$$\begin{aligned}&\inf_{0\leq\|g\|_2^2\leq 1}\int_0^1 (1-g(\theta))^2 w(\theta)+\lambda g(\theta)^2)d\theta\\ \leq&\inf_{0\leq\|g\|_2^2\leq 1}\int_0^1 (1-g(\theta))^2 v(\theta)+\lambda g(\theta)^2)d\theta \qquad (\#)\end{aligned}$$

We can further simplify the infimums in (#) by explicitly evaluating them:

Lemma 12

$$\inf_{0\leq\|g\|_2^2\leq 1}\int_0^1 (1-g(\theta))^2 w(\theta)+\lambda g(\theta)^2)d\theta = \int_0^1 (\lambda - \frac{\lambda^2}{w(\theta)+\lambda})d\theta$$

To evaluate the l.h.s. infimum it turns out the obvious thing works: namely, minimizing the integrand, $(1-g(\theta))^2 w(\theta)+\lambda g(\theta)^2)$.
To tackle this problem, let's try to first minimize it pointwise, by fixing an arbitrary point $\theta \in [0,1]$ and considering this as a function of the new variable $y = g(\theta)$, where g varies over all functions that have the above constraint.
Then since w is a fixed function, $w(\theta) = w$ is a constant, and we obtain the function $L(y) = (1-y)^2 w + \lambda y^2$. L is just a quadratic, hence a differentiable function, in y, and so its minimum is obtained by elementary calculus methods. We take the derivative $L'(y)$ and set $L'(y) = 0$:

$$\begin{aligned}&L'(y) = -2(1-y)w + 2\lambda y = 0\\ \Rightarrow&(1-y)w = \lambda y\\ \Rightarrow&w = y(\lambda + w)\\ \Rightarrow&y = \frac{w}{\lambda + w} \quad \text{which is well-defined since } \lambda > 0, w \geq 0\end{aligned}$$

So we obtain the function $g(x) = \frac{w(x)}{\lambda + w(x)}$. We now have two things to check:

1) that this particular function g is a candidate for giving the minimum value of $(1 - g(\theta))^2 w(\theta) + \lambda g(\theta)^2)$; i.e. this g must satisfy the inf constraint: $0 \leq \|g\|_2^2 \leq 1$.

2) that this minimum value of the integrand does give us the desired inequality, $\int_0^1 \frac{1}{\lambda + w(\theta)} d(\theta) \geq \int_0^1 \frac{1}{\lambda + v(\theta)} d(\theta)$, of the proposition.

1) Writing w again as a function (not just a particular value as before), we have

$$0 \leq \frac{w}{\lambda + w} \leq \frac{\lambda + w}{\lambda + w} = 1, \text{ since } \lambda \neq 0;$$

$$\Rightarrow 0 \leq (\frac{w}{\lambda + w})^2 \leq 1$$
$$\Rightarrow 0 \leq \int_0^1 (\frac{w(\theta)}{\lambda + w(\theta)})^2 d\theta \leq 1$$
$$\text{i.e. } 0 \leq \|\frac{w}{\lambda + w}\|_2^2 \leq 1$$

When the integrand is evaluated at $g = \frac{w}{\lambda+w}$, we obtain

$$\begin{aligned}
(1-\frac{w}{\lambda+w})^2 w + \lambda(\frac{w}{\lambda+w})^2 &= \frac{\lambda^2 w}{(\lambda+w)^2} + \frac{\lambda w^2}{(\lambda+w)^2} \\
&= \frac{\lambda w(\lambda+w)}{(\lambda+w)^2} \\
&= \lambda\frac{w}{\lambda+w} \\
&= \lambda\frac{w+\lambda-\lambda}{\lambda+w} \\
&= \lambda\frac{w+\lambda}{\lambda+w} - \frac{\lambda^2}{\lambda+w} \\
&= \lambda - \frac{\lambda^2}{\lambda+w}
\end{aligned}$$

Thus, the function $g = \frac{w}{\lambda+w}$ minimizes the integrand $(1-g)^2 w + \lambda g^2$. Therefore (#) becomes

$$\int_0^1 (\lambda - \frac{\lambda^2}{w(\theta)+\lambda})d\theta \leq \int_0^1 (\lambda - \frac{\lambda^2}{v(\theta)+\lambda})d\theta$$

2) Continuing from the previous line, we obtain

$$\begin{aligned}
\int_0^1 \lambda d\theta - \lambda^2 \int_0^1 \frac{1}{w(\theta)+\lambda}d\theta &\leq \int_0^1 \lambda d\theta - \lambda^2 \int_0^1 \frac{1}{v(\theta)+\lambda}d\theta \\
\Leftrightarrow \lambda - \lambda^2 \int_0^1 \frac{1}{w(\theta)+\lambda}d\theta &\leq \lambda - \lambda^2 \int_0^1 \frac{1}{v(\theta)+\lambda}d\theta \\
\Leftrightarrow -\lambda^2 \int_0^1 \frac{1}{w(\theta)+\lambda}d\theta &\leq -\lambda^2 \int_0^1 \frac{1}{v(\theta)+\lambda}d\theta \\
\Leftrightarrow \int_0^1 \frac{1}{w(\theta)+\lambda}d\theta &\geq \int_0^1 \frac{1}{v(\theta)+\lambda}d\theta
\end{aligned}$$

Since $\lambda \geq 0$ was arbitrary, this proves the first implication of the proposition.

($\Leftarrow$) Assume that $\int_0^1 \frac{1}{w(\theta)+\lambda} d\theta \geq \int_0^1 \frac{1}{v(\theta)+\lambda} d\theta$ for any $\lambda \geq 0$. After we applied the T-transform in the first part above, all subsequent calculations were equivalencies, and so here we get back that same first T-transform by the preceding lemma:

$$\int_0^1 \frac{1}{w(\theta)+\lambda} d\theta \geq \int_0^1 \frac{1}{v(\theta)+\lambda} d\theta$$

$$\Rightarrow \inf_{0 \leq t \leq 1} (K_w(t) + \lambda t) \leq \inf_{0 \leq t \leq 1} (K_v(t) + \lambda t)$$

Now apply the t-transform (which we have seen inverts the T-transform) to each side of this equation to get back the original functions K_w,K_v:

$$\Rightarrow \sup_{\lambda \geq 0} [\inf_{0 \leq t \leq 1} (K_w(t) + \lambda t) - \lambda x] \leq \sup_{\lambda \geq 0} [\inf_{0 \leq t \leq 1} (K_v(t) + \lambda t) - \lambda x]$$

$$\Rightarrow K_w(x) \leq K_v(x) \text{ for any } x \in [0, 1]$$

This proves the other implication of the proposition. □

Remark: One can proceed to prove this theorem by using the regular convex conjugate transform, along the same lines of the proof we gave using the convex conjugate transform of the function F_w in Section 3.

References

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics* , Springer-Verlag, New York, (1989)

[2] G. B. Folland, *Real Analysis - Modern Techniques and their Applications* , Wiley-Interscience, New York, (1984)

[3] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* , Cambridge University Press, Cambridge [England]; New York, (1988)

[4] G. H. Hardy, J. E. Littlewood, and G. Pólya, Some Simple Inequalities Satisfied by Convex Functions, *Messenger of Math* **58**: 145-152 (1929)

[5] I. Klemes, Finite Toeplitz Matrices and Sharp Littlewood Conjectures, *St. Petersburg Math. J.* **13**: 27-40 (2002)

[6] A. W. Marshall and I. Olkin, *Inequalities: theory of majorization and its applications* , Academic Press, New York, (1979)

[7] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I* , Springer-Verlag, Berlin; New York, (1970)