The Lambert W Function in Optimization

Jonathan M. Borwein and Scott B. Lindstrom

CARMA, University of Newcastle

Wayne State, April 12, 2016
OVA7, Alicante June 3, 2016
Optimization Down Under, July 18-22, 2016

Last Revised April 15, 2016

https://www.carma.newcastle.edu.au/jon/WinOpt.pdf

Meetings with the Lambert W function and other special functions in optimisation and analysis

Jonathan Borwein
Laureate Professor and Director
CARMA
www.carma.newcastle.edu.au

Revised 24-05-2016

"I never run for trains." Nasim Nicholas Taleb (The Black Swan)

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

2016 Presentations

Jonathan Borwein FRSC FAA FAAA
Laureate Professor
and Director CARMA
www.carma.newcastle.edu.au

2016 Presentations as
Distinguished Scholar in Residence
Western University, London Ontario

Western UNIVERSITY · CANADA

April 12-13 : Owens Lectures
Wayne State University
1. Lambert W in Optimization
2. Walking on Numbers

CARMA

WAYNE STATE

"I never run for trains." Nasim Nicholas Taleb
(The Black Swan)

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Revised 4-02-2016

Jon's website:

https://www.carma.newcastle.edu.au/jon/WinOpt.pdf

CARMA

Computer Assisted Research Mathematics and its Applications (CARMA) Priority Research Centre

https://carma.newcastle.edu.au/

Scott's website:

https://carma.newcastle.edu.au/findsem.php?n=395

Outline I

Outline II

- Occurrences in Composition
- Occurrences in Infimal Convolution
- Occurrences in Homotopy

5. Homotopy and Entropy Solutions of Inverse Problems
   - An Optimization Problem
   - A General Implementation
   - Computed Examples

6. Conclusion
   - Further Merits of SCAT and CCAT
   - Bibliography

Definition

- Lambert W is the inverse of $x \mapsto x\exp(x)$. The real inverse is two-valued, as shown in Figure 1.

Definition

- Lambert W is the inverse of $x \mapsto x\exp(x)$. The real inverse is two-valued, as shown in Figure 1.
- We are interested in the principal branch with Taylor series

$$W(x) = \sum_{k=1}^{\infty} \frac{(-k)^{k-1}}{k!} x^k$$

with radius of convergence $1/e$.

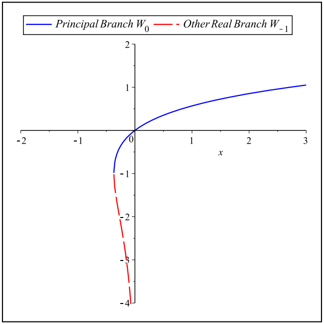


Figure: The real branches of the Lambert W function.

Basic Properties

1. Implicit differentiation leads to
$$W'(x) = \frac{W(x)}{x\,(1 + W(x))}.$$
2. W is concave on $(-1/e, \infty)$ and positive on $(0, \infty)$.
3. $(\log \circ W)(z) = \log(z) - W(z)$ is concave; since W is log concave on $(0, \infty)$.
4. $\exp(W(z)) = z/W(z)$ is concave.

The Power of Naming

> *Besides it is an error to believe that rigor in the proof is the enemy of simplicity. – David Hilbert*

The Power of Naming

> *Besides it is an error to believe that rigor in the proof is the enemy of simplicity. – David Hilbert*

- W is an excellent counter-example to Stigler's Law of Eponymy (which asserts that an idea is always named after the *last* person to discover it).

Figure: Johann Heinrich Lambert (1728–1777)

Trefethen's Problem

In **2002** Nick Trefethen published ten numerical challenge problems in *SIAM Review* [3]. Several are in optimization.

Trefethen's Problem

In **2002** Nick Trefethen published ten numerical challenge problems in *SIAM Review* [3]. Several are in optimization.

Example (Trefethen's ninth problem [3])

The problem is posed as follows.

The integral

$$I(\alpha) = \int_0^2 [2 + \sin(10\alpha)] x^\alpha \sin\left(\frac{\alpha}{2-x}\right) \mathrm{d}x$$

depends on the parameter α. *What is the value* $\alpha \in [0, 5]$ *at which* $I(\alpha)$ *achieves its maximum?*

The Computer Informs the Scientist

- What *we* know about a function often matters less than what our CAS (say *Maple*, *Mathematica*, or *SAGE*) does.

The Computer Informs the Scientist

- What *we* know about a function often matters less than what our CAS (say *Maple*, *Mathematica*, or *SAGE*) does.
- At left: what *Maple* knows about Meijer-G.

MeijerG - Meijer G function

Calling Sequence

```
MeijerG([as, bs], [cs, ds], z)
```

Parameters

as	-	list of the form [a1, ..., am]; first group of numerator Γ parameters
bs	-	list of the form [b1, ..., bn]; first group of denominator Γ parameters
cs	-	list of the form [c1, ..., cp]; second group of numerator Γ parameters
ds	-	list of the form [d1, ..., dq]; second group of denominator Γ parameters
z	-	expression

Description

- The Meijer G function is defined by the inverse Laplace transform

$$\mathrm{MeijerG}([as, bs], [cs, ds], z) = \frac{1}{2\pi I}\oint_L \frac{\Gamma(1-as+y)\,\Gamma(cs-y)}{\Gamma(bs-y)\,\Gamma(1-ds+y)} z^y \mathrm{d}y$$

where

$$as = [a1, ..., am],\ \Gamma(1-as+y) = \Gamma(1-a1+y)\ ...\ \Gamma(1-am+y)$$

$$bs = [b1, ..., bn],\ \Gamma(bs-y) = \Gamma(b1-y)\ ...\ \Gamma(bn-y)$$

$$cs = [c1, ..., cp],\ \Gamma(cs-y) = \Gamma(c1-y)\ ...\ \Gamma(cp-y)$$

$$ds = [d1, ..., dq],\ \Gamma(1-ds+y) = \Gamma(1-d1+y)\ ...\ \Gamma(1-dq+y)$$

and L is one of three types of integration paths $L_{\gamma+\infty I}$, L_{∞}, and $L_{-\infty}$.

Contour L_{∞} starts at $\infty + I\,\phi 1$ and finishes at $\infty + I\,\phi 2\ (\phi 1 < \phi 2)$.

Contour $L_{-\infty}$ starts at $-\infty + I\,\phi 1$ and finishes at $-\infty + I\,\phi 2\ (\phi 1 < \phi 2)$.

Contour $L_{\gamma+\infty I}$ starts at $\gamma - \infty I$ and finishes at $\gamma + \infty I$.

A Solution to Trefethen's Problem

- $I(\alpha)$ is expressible in terms of a *Meijer-G* function.
- Unlike most humans, *Mathematica* and *Maple* will figure this out.
  - Help files or a web search then inform the scientist.
  - This is a measure of the changing environment.
- Below: the exact form of $I(\alpha)$ as given by *Maple*.

$$I(\alpha) = 4\sqrt{\pi}\,\Gamma(\alpha)\,G^{3,0}_{2,4}\left(\frac{\alpha^2}{16}\,\middle|\,\begin{matrix}\frac{\alpha+2}{2},\ \frac{\alpha+3}{2}\\ \frac{1}{2},\ \frac{1}{2},\ 1,\ 0\end{matrix}\right)[\sin(10\,\alpha)+2].$$

Short Random Walks

- Assuming the Meier-G function is well implemented, one can now use any good numerical optimiser.
- The Meijer-G function has also been instrumental in producing new results on a hundred-year-old topic:

Short Random Walks

- Assuming the Meier-G function is well implemented, one can now use any good numerical optimiser.
- The Meijer-G function has also been instrumental in producing new results on a hundred-year-old topic:

Example (Moments of random walks [10])

The moment function of an $n-$step random walk in the plane is:

$$M_n(s) = \int_{[0,1]^n} \left| \sum_{k=1}^{n} e^{2\pi x_k i} \right|^s \mathrm{d}(x_1, \ldots, x_{n-1}, x_n).$$

A Moment Function

The first breakthrough in [10] makes use of Meijer-G:

Theorem (Meijer-G form for M_3)

For s not an odd integer,

$$M_3(s) = \frac{\Gamma(1+\frac{s}{2})}{\sqrt{\pi}\,\Gamma(-\frac{s}{2})}\, G^{21}_{33}\left(\begin{matrix} 1,1,1 \\ \frac{1}{2}, -\frac{s}{2}, -\frac{s}{2} \end{matrix} \middle| \frac{1}{4} \right). \tag{1}$$

A Moment Function

The first breakthrough in [10] makes use of Meijer-G:

Theorem (Meijer-G form for M_3)

For s not an odd integer,

$$M_3(s) = \frac{\Gamma(1+\frac{s}{2})}{\sqrt{\pi}\,\Gamma(-\frac{s}{2})}\, G^{21}_{33}\left(\begin{matrix} 1,1,1 \\ \frac{1}{2},-\frac{s}{2},-\frac{s}{2} \end{matrix}\,\middle|\,\frac{1}{4}\right). \tag{1}$$

- Equation (1) was first found by Crandall via CAS and proven in [10] using residue calculus methods.
- $M_3(s)$ is among the first non-trivial higher order Meijer-G functions to be placed in closed form. (Also $M_4(s)$.)

A New Result on an Old Topic

Theorem (Meijer-G form for M_4)

For $\Re s > -2$ and s not an odd integer

$$M_4(s) = \frac{2^s}{\pi} \frac{\Gamma(1+\frac{s}{2})}{\Gamma(-\frac{s}{2})} G_{44}^{22}\left(\begin{matrix} 1, \frac{1-s}{2}, 1, 1 \\ \frac{1}{2} - \frac{s}{2}, -\frac{s}{2}, -\frac{s}{2}, -\frac{s}{2} \end{matrix} \middle| 1 \right). \tag{2}$$

A New Result on an Old Topic

Theorem (Meijer-G form for M_4)

For $\Re s > -2$ and s not an odd integer

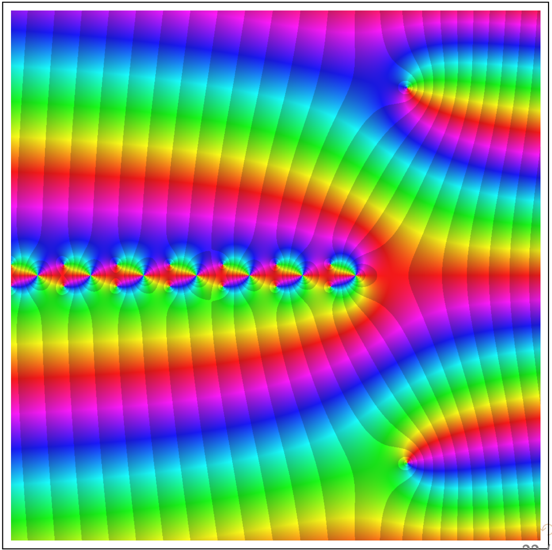
$$M_4(s) = \frac{2^s}{\pi} \frac{\Gamma(1+\frac{s}{2})}{\Gamma(-\frac{s}{2})} G^{22}_{44}\left(\begin{matrix} 1, \frac{1-s}{2}, 1, 1 \\ \frac{1}{2} - \frac{s}{2}, -\frac{s}{2}, -\frac{s}{2}, -\frac{s}{2} \end{matrix} \middle| 1 \right). \tag{2}$$

This, together with the first result, led to useful results, including:

Closed hypergeometric form for the radial density of a 3-step walk:

$$p_3(\alpha) = \frac{2\sqrt{3}\alpha}{\pi\,(3+\alpha^2)}\, {}_2F_1\left(\begin{matrix} \frac{1}{3}, \frac{2}{3} \\ 1 \end{matrix} \middle| \frac{\alpha^2\left(9-\alpha^2\right)^2}{\left(3+\alpha^2\right)^3} \right) \tag{3}$$

The moment function M_4 drawn from (2) in the Calendar Complex Beauties 2016.



Knuth's Series Problem

We continue with an account of the solution in [5], to a problem posed by Donald E. Knuth in the November 2000 issues of the *American Mathematical Monthly*.

Problem 10832

Evaluate

$$S = \sum_{k=1}^{\infty}\left(\frac{k^k}{k!e^k} - \frac{1}{\sqrt{2\pi k}}\right).$$

See [18] for the published solution.

A Numerical Solution

Problem 10832

$$S = \sum_{k=1}^{\infty} \left(\frac{k^k}{k! e^k} - \frac{1}{\sqrt{2\pi k}} \right).$$

Maple produced the approximation

$$S \approx -0.084069508727655996461.$$

With "Smart Lookup" feature, the Inverse Symbolic Calculator* yielded:

$$S \approx -\frac{2}{3} - \frac{1}{\sqrt{2\pi}} \zeta\left(\frac{1}{2}\right). \tag{4}$$

Available at `http://isc.carma.newcastle.edu.au/`

A CAS Solution

- Calculations to higher precision (50 decimal digits) confirmed this approximation. Are we done?
- Why would such an identity hold?

A CAS Solution

- Calculations to higher precision (50 decimal digits) confirmed this approximation. Are we done?
- Why would such an identity hold?
  - One clue was the speed with which *Maple* calculated the precise value of this slowly convergent sum. *Maple* clearly knew something we did not ...

A CAS Solution

- Calculations to higher precision (50 decimal digits) confirmed this approximation. Are we done?
- Why would such an identity hold?
  - One clue was the speed with which *Maple* calculated the precise value of this slowly convergent sum. *Maple* clearly knew something we did not ...
  - We discovered *Maple* was using the Lambert W function.

A CAS Solution

- Calculations to higher precision (50 decimal digits) confirmed this approximation. Are we done?
- Why would such an identity hold?
  - One clue was the speed with which *Maple* calculated the precise value of this slowly convergent sum. *Maple* clearly knew something we did not ...
  - We discovered *Maple* was using the Lambert W function.
- Another clue was the appearance of $\zeta(1/2)$ in the above experimental identity, together with an obvious allusion to Stirling's formula in the original problem.

A Conjectured Identity

We Conjectured the Identity

$$\sum_{k=1}^{\infty}\left(\frac{1}{\sqrt{2\pi k}}-\frac{P(1/2,k-1)}{(k-1)!\sqrt{2}}\right)=\frac{1}{\sqrt{2\pi}}\zeta\left(\frac{1}{2}\right)$$

- Here $P(x,n)$ denotes the *Pochhammer symbol* $x(x+1)\cdots(x+n-1)$, and the binomial coefficients on the left hand side are the same as those of the function $1/\sqrt{2-2x}$.
- *Maple* successfully evaluated this summation as shown on the right hand side.

We now needed to establish that

$$\sum_{k=1}^{\infty}\left(\frac{k^k}{k!e^k}-\frac{P(1/2,k-1)}{(k-1)!\sqrt{2}}\right) = -\frac{2}{3}.$$

Guided by the presence of the Lambert W function,

$$W(z) = \sum_{k=1}^{\infty}\frac{(-k)^{k-1}z^k}{k!},$$

an appeal to Abel's limit theorem suggested

$$\lim_{z\to 1}\left(\frac{dW(-z/e)}{dz}+\frac{1}{\sqrt{2-2z}}\right)=\frac{2}{3}.$$

Maple was able to evaluate this limit and so establish the identity.

Proving the Identity

The identity relies on the following reversion [16]. Let $p = \sqrt{2(1+ez)}$ with $z = We^W$, so that

$$\frac{p^2}{2} - 1 = W\exp(1+W) = -1 + \sum_{k\geq 1}\left(\frac{1}{k!} - \frac{1}{(k-1)!}\right)(1+W)^k$$

and revert to $1 + W = p - \frac{p^2}{3} + \frac{11}{72}p^3 + \ldots$ for $|p| < \sqrt{2}$.

This combines with $W'(x) = \frac{W(x)}{x\,(1+W(x))}$ to prove the identity.

Remark on Generalisation

Proposition. ($\zeta(s)$ for $0 < s < \infty, s \neq 1$)

For $0 < \mathrm{Re}\ s < 1$ in the complex plane,

$$\sum_{k=1}^{\infty} \left(\frac{1}{k^s} - \frac{\Gamma(k-s)}{\Gamma(k)} \right) = \zeta(s). \tag{5}$$

Remark on Generalisation

Proposition. ($\zeta(s)$ for $0 < s < \infty, s \neq 1$)

For $0 < \text{Re } s < 1$ in the complex plane,

$$\sum_{k=1}^{\infty} \left(\frac{1}{k^s} - \frac{\Gamma(k-s)}{\Gamma(k)} \right) = \zeta(s). \quad (5)$$

Now *Maple*'s summation tools can reduce this to

$$\sum_{k=1}^{N} \frac{1}{k^s} - \frac{\Gamma(N+1-s)}{(1-s)\Gamma(N)} \to \zeta(s). \quad (6)$$

For any given rational $s \in (0, \infty)$ *Maple* will evaluate the limit by the Euler-Maclaurin method.

Remark on Generalisation

Now *Maple*'s summation tools can reduce this to

$$\sum_{k=1}^{N} \frac{1}{k^s} - \frac{\Gamma(N+1-s)}{(1-s)\Gamma(N)} \to \zeta(s). \tag{7}$$

For any given rational $s \in (0, \infty)$ *Maple* will evaluate the limit by the Euler-Maclaurin method. Consulting the DLMF* we discover

$$\zeta(s) = \sum_{k=1}^{N} \frac{1}{k^s} + \frac{N^{1-s}}{s-1} - s\int_N^\infty \frac{x - \lfloor x \rfloor}{x^{s+1}}\,\mathrm{d}x.$$

*Found at http://www.dlmf.gov.

Remark on Generalisation

For any given rational $s \in (0, \infty)$ *Maple* will evaluate the limit by the Euler-Maclaurin method. Consulting the DLMF* we discover

$$\zeta(s) = \sum_{k=1}^{N} \frac{1}{k^s} + \frac{N^{1-s}}{s-1} - s \int_N^{\infty} \frac{x - \lfloor x \rfloor}{x^{s+1}} \, \mathrm{d}x.$$

Since the integral tends to zero for $s > 0$ and

$$\lim_{N \to \infty} \frac{\Gamma(N+1-s)}{(1-s)\Gamma(N)} - \frac{N^{1-s}}{1-s} = 0,$$

we can also produce an explicit human proof.

*Found at http://www.dlmf.gov

An Open Question

Can one find a solution for general $s \neq \frac{1}{2} \in (0,1)$?

Based on (5) and the Stirling approximation for $\Gamma(k+s) \approx \sqrt{2\pi}\, e^{-k} k^{k+s-1/2}$ we obtain

$$\sum_{k=1}^{\infty} \left(\frac{1}{\sqrt{2\pi}\, k^s} - \frac{k^{k+1/2-s}}{k!\, e^k} \right) - \frac{\zeta(s)}{\sqrt{2\pi}} = \kappa(s)\,. \tag{8}$$

We have $\kappa(1/2) = 2/3$, but it remains to evaluate $\kappa(s) \in \mathbb{R}$ more generally. Our question is closely allied to that of asking if

$$W_s(x) = \sum_{k=1}^{\infty} \frac{k^{k+1/2-s}}{k!} x^k \tag{9}$$

for $s \neq 1/2$ can be analysed in terms of W.

A Plot of the Function in Question

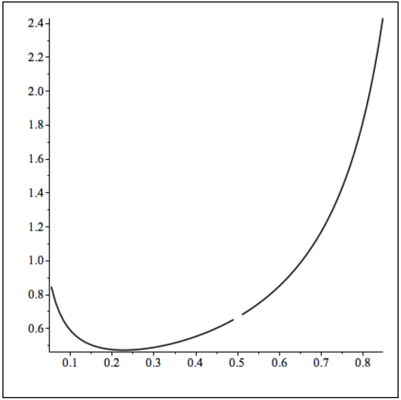


Figure: The function κ to the left and right of $s = 1/2$.

Definition of Convex Conjugate

For a function $f: X \to [-\infty, \infty]$ the convex conjugate is

the function $f^*: X^* \to [-\infty, \infty]$ given by

$$f^*(y) = \sup_{x \in X} \langle y, x \rangle - f(x). \tag{10}$$

Here X is a Euclidean, Hilbert, or Banach space.

Definition of Convex Conjugate

For a function $f: X \to [-\infty, \infty]$ the convex conjugate is the function $f^*: X^* \to [-\infty, \infty]$ given by

$$f^*(y) = \sup_{x \in X} \langle y, x \rangle - f(x). \tag{10}$$

Here X is a Euclidean, Hilbert, or Banach space.

- The function f^* is always convex (if possibly always infinite).
- If f is lower semicontinuous, convex, proper, $(f^*)^* = f$.
- In particular if we show (by CAS) a function $g = f^*$ for some alert$f$, then g is necessarily convex.

Visualizing Convex Conjugates

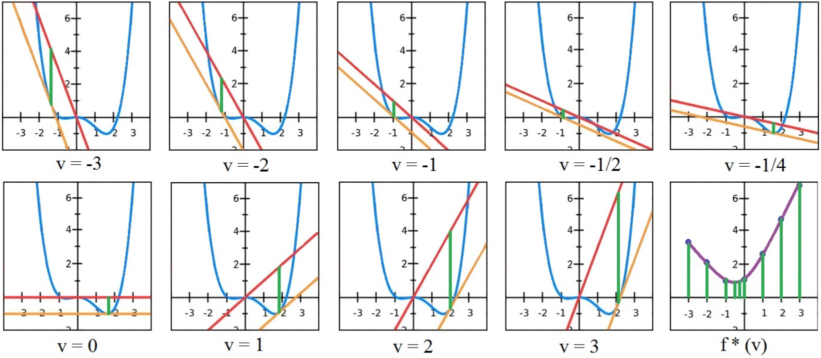


Figure: The construction of f^* is shown for a blue function f. The inputs of f^* may be thought of as slopes of the lines through the origin. For each input, we obtain the corresponding output by taking a parallel line and sliding it down as far away from the original line as as it can go while still touching the curve of the function f. The output is the vertical distance between the two lines [19]

Computing a Closed Form

Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$.
Then

$$f^*(y) = \sup_{x\in\mathbb{R}}\{\langle y, x\rangle - f(x)\}$$
$$= \sup_{x\in\mathbb{R}}\{yx - x^2\}.$$

Computing a Closed Form

Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$.
Then

$$\begin{aligned} f^*(y) &= \sup_{x\in\mathbb{R}}\{\langle y, x\rangle - f(x)\} \\ &= \sup_{x\in\mathbb{R}}\{yx - x^2\}. \end{aligned}$$

- Differentiating $yx - x^2$ and using $y - 2x = 0$, we find that $yx - x^2$ attains its supremum when $x = \frac{y}{2}$.

Computing a Closed Form

Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then

$$f^*(y) = \sup_{x\in\mathbb{R}}\{\langle y, x\rangle - f(x)\} = \sup_{x\in\mathbb{R}}\{yx - x^2\}.$$

- Differentiating $yx - x^2$ and using $y - 2x = 0$, we find that $yx - x^2$ attains its supremum when $x = \frac{y}{2}$.

We substitute to obtain $f^*(y) = y(\frac{y}{2}) - (\frac{y}{2})^2 = \frac{y^2}{4}$.

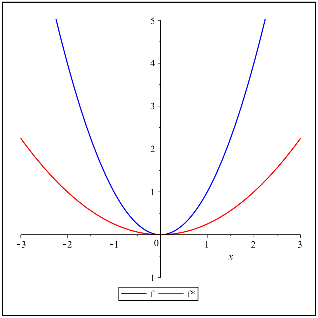


Figure: The function $f(x) = x^2$ and its conjugate $f^*(y) = \frac{y^2}{4}$ [19].

Some Important Examples

- For $1/p + 1/q = 1$ with $p, q > 1$,

$$\left(\frac{|\cdot|^p}{p}\right)^* = \frac{|\cdot|^q}{q}.$$

- The *energy* function $\frac{|\cdot|^2}{2}$ is the only self-conjugate function.
- The *log barrier* $f(x) = -\log x$ for $x > 0$ has conjugate conjugate $f^*(y) = -1 - \log y$ for $x < 0$.
- The Boltzmann-Shannon entropy $y \log(y) - y$ is the convex conjugate of $\exp(x)$ (and vice-versa since $\exp(x)$ is convex).

Our *Maple* packages SCAT & CCAT [7] automate all this and more subtle ideas such as iterated conjugation: see http://carma.newcastle.edu.au/ConvexFunctions/SCAT.ZIP.

Addition and Convolution

The convex conjugate exchanges addition of functions with their infimal convolution

$$(f \,\square\, g)(y) = \inf_{x \in X} f(y-x) + g(x).$$

Indeed $(f \,\square\, g)^* = f^* + g^*$ always holds and under mild hypotheses

$$(f+g)^* = f^* \,\square\, g^*.$$

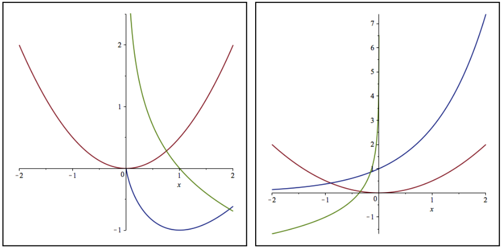


Figure: The energy, log barrier and negative entropy (L) and duals (R).

Variable Separability

Suppose f is *variable separable*. That is to say that

$$f(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} f_j(x_j)$$

where each f_j is convex. Then f is convex and

$$f^*(y_1, y_2, \ldots y_n) = \sum_{j=1}^{n} f_j^*(y_j).$$

From such building blocks, and the *Fenchel duality* theorem – for $f + g \circ A$ – [8] or Theorem 5 below, many other convex conjugates engaging W are accessible.

Matrix Functions

It is also possible to induce functions of matrices as follows.

Matrix Functions

It is also possible to induce functions of matrices as follows.

- Let f be a symmetric proper and lower semicontinuous convex function of n variables, and let A be a symmetric matrix with real spectrum $\lambda(A)$.
- Then

$$\widehat{f}(A) = f(\lambda(A))$$

  induces a proper and lower semicontinuous convex matrix function.

Matrix Functions

It is also possible to induce functions of matrices as follows.

- Let f be a symmetric proper and lower semicontinuous convex function of n variables, and let A be a symmetric matrix with real spectrum $\lambda(A)$.
- Then
$$\widehat{f}(A) = f(\lambda(A))$$
induces a proper and lower semicontinuous convex matrix function.
- So $f(x) = -\sum_{k=1}^{n} \log(x_k)$ induces $\widehat{f}(A) = -\log(\det(A))$.

Matrix Functions

It is also possible to induce functions of matrices as follows.

- Let f be a symmetric proper and lower semicontinuous convex function of n variables, and let A be a symmetric matrix with real spectrum $\lambda(A)$.
- Then
$$\widehat{f}(A) = f(\lambda(A))$$
induces a proper and lower semicontinuous convex matrix function.
- So $f(x) = -\sum_{k=1}^{n} \log(x_k)$ induces $\widehat{f}(A) = -\log(\det(A))$.

Moreover,
$$\widehat{f^*}(A) = \left(\widehat{f}(A)\right)^*.$$

A Note on Closed Forms

- The notion of a closed form for a given function is an always-changing issue.
  - While $x \exp x$ is elementary $W(x)$ is not, since arbitrary inversion is not permitted in the definition of *elementary*.

A Note on Closed Forms

- The notion of a closed form for a given function is an always-changing issue.
  - While $x \exp x$ is elementary $W(x)$ is not, since arbitrary inversion is not permitted in the definition of *elementary*.
- We consider a closed form roughly to be a form which is finitary and computationally effective. See, for example, [6] available at `https://www.carma.newcastle.edu.au/jon/closed-form.pdf`.
- Once a computationally effective closed form is available, all of classical convex duality theory is accessible.

Log Convex Functions

Definition

A log convex function g is a positive function such that $f = \log g$ is convex. Thence

$$g(x) = e^{f(x)}.$$

Log Convex Functions

Definition

A log convex function g is a positive function such that $f = \log g$ is convex. Thence

$$g(x) = e^{f(x)}.$$

- log convexity – useful in statistics – may be thought of as a strengthening of convexity and is implied by $1/g > 0$ being concave.
- We are interested in the convex conjugates of such functions:

$$g^*(y) = \sup_{x \in X} \{yx - e^{f(x)}\}$$

Convex Conjugates of Log Convex Functions

Given the function f, we may solve, as before, by taking the derivative and setting it equal to zero to obtain

$$y = f'(x)e^{f(x)}.$$

Convex Conjugates of Log Convex Functions

Given the function f, we may solve, as before, by taking the derivative and setting it equal to zero to obtain

$$y = f'(x)e^{f(x)}.$$

If we can solve this equation for $x = s(y)$, we can express the conjugate in closed form as

$$g^*(y) = y \cdot s(y) - g(s(y)).$$

We will explore a useful class of functions for which W shows up quite naturally in their closed forms.

Two Examples with W

Our *Maple* package *SCAT* provides two such examples which we can also easily verify by the methods above.

- For $g(x) = e^{e^x}$, we have

$$g^*(y) = \begin{cases} y\left(\log(y) - W(y) - \frac{1}{W(y)}\right) & \text{if } y > 0 \\ -1 & \text{if } y = 0 \\ \infty & \text{if } y < 0 \end{cases}.$$

Two Examples with W

Our *Maple* package *SCAT* provides two such examples which we can also easily verify by the methods above.

- For $g(x) = e^{e^x}$, we have

$$g^*(y) = \begin{cases} y\left(\log(y) - W(y) - \frac{1}{W(y)}\right) & \text{if } y > 0 \\ -1 & \text{if } y = 0 \\ \infty & \text{if } y < 0 \end{cases}.$$

- For $g(x) = e^{\frac{x^2}{2}}$, we have

$$g^*(y) = |y|\left(\sqrt{W(y^2)} - \frac{1}{\sqrt{W(y^2)}}\right) \text{ for all } y.$$

Seeking a General Closed Form

If we can first solve the equation

$$f'(x)^{\alpha+1} = \gamma f(x) \tag{11}$$

for some α and nonzero γ, we will be able to express g^* in closed form using W.

Seeking a General Closed Form

If we can first solve the equation

$$f'(x)^{\alpha+1} = \gamma f(x) \tag{11}$$

for some α and nonzero γ, we will be able to express g^* in closed form using W. Indeed, since

$$y = f'(x)e^{f(x)},$$

we raise both sides to the $\alpha + 1$ power to obtain

$$y^{\alpha+1} = f'(x)^{\alpha+1}e^{(\alpha+1)f(x)} = \gamma f(x)e^{(\alpha+1)f(x)}.$$

Seeking a General Closed Form

If we can first solve the equation

$$f'(x)^{\alpha+1} = \gamma f(x) \tag{11}$$

for some α and nonzero γ, we will be able to express g^* in closed form using W. Indeed, since

$$y = f'(x)e^{f(x)},$$

we raise both sides to the $\alpha + 1$ power to obtain

$$y^{\alpha+1} = f'(x)^{\alpha+1}e^{(\alpha+1)f(x)} = \gamma f(x)e^{(\alpha+1)f(x)}.$$

Finally, we multiply both sides by $\frac{\alpha+1}{\gamma}$ and can use W to write:

$$(\alpha + 1)f(x) = W\left((\alpha + 1)\frac{y^{\alpha+1}}{\gamma}\right).$$

We obtain the following closed forms for $(\exp \circ f)^*(y)$: we write

$$f(x) = \frac{W\left((\alpha+1)\frac{y^{\alpha+1}}{\gamma}\right)}{\alpha+1}$$

$$x = b\left(\frac{W\left((\alpha+1)\frac{y^{\alpha+1}}{\gamma}\right)}{\alpha+1}, y\right)$$

Here $b(x,y) = f^{-1}(x)$ in the invertible case and $b(x,y)$ is the pre-image choice in $f^{-1}(x)$ such that $x \cdot y$ is maximized otherwise.

We obtain the following closed forms for $(\exp \circ f)^*(y)$: we write

$$f(x) = \frac{W\left((\alpha+1)\frac{y^{\alpha+1}}{\gamma}\right)}{\alpha+1}$$

$$x = b\left(\frac{W\left((\alpha+1)\frac{y^{\alpha+1}}{\gamma}\right)}{\alpha+1}, y\right)$$

Here $b(x,y) = f^{-1}(x)$ in the invertible case and $b(x,y)$ is the pre-image choice in $f^{-1}(x)$ such that $x \cdot y$ is maximized otherwise.

These yield the closed form for g^*:

$$g^*(y) = y \cdot b(d(y), y) - \exp(d(y)) \text{ where}$$

$$d(y) = \frac{W\left((\alpha+1)\frac{y^{\alpha+1}}{\gamma}\right)}{\alpha+1}.$$

Comparing to our Previous Examples

We can see quite nicely how this relates to our previous examples.

- In the case of our example $g(x) = \exp(\exp(x))$, we have

$$b(x, y) = f^{-1}(x) = \log(x).$$

Comparing to our Previous Examples

We can see quite nicely how this relates to our previous examples.

- In the case of our example $g(x) = \exp(\exp(x))$, we have

$$b(x,y) = f^{-1}(x) = \log(x).$$

- In the case of our example $g(x) = \exp\left(\frac{|x|^p}{p}\right)$, we have

$$b(x,y) = \begin{cases} (p\cdot x)^{\frac{1}{p}} & \text{if } y \geq 0 \\ -(p\cdot x)^{\frac{1}{p}} & \text{if } y < 0 \end{cases}. \tag{12}$$

A Simplified General Form

Using the fact that $\exp(W(x)) = x/W(x)$, we can further simplify the expression of our general closed form to:

Closed form when (11) holds

$$g^*(y) = y \cdot b\left(\frac{W\left((\alpha+1)\frac{y^{\alpha+1}}{\gamma}\right)}{\alpha+1}, y\right) - \left(\frac{(\alpha+1)\frac{y^{\alpha+1}}{\gamma}}{W\left((\alpha+1)\frac{y^{\alpha+1}}{\gamma}\right)}\right)^{\frac{1}{\alpha+1}}.$$

While this isn't especially nice to look at, it simplifies greatly for certain choices of f (ergo, choices of γ, α). More importantly, it is very clean from a computational standpoint.

A Class of Functions

- Since our use of W relies upon being able to solve

$$f'(x)^{\alpha+1} = \gamma f(x),$$

we ask for what kind of function f this is possible.

A Class of Functions

- Since our use of W relies upon being able to solve
$$f'(x)^{\alpha+1} = \gamma f(x),$$
we ask for what kind of function f this is possible.

With initial condition $f(0) = \beta$, *Maple*'s built-in ODE solver returns

$$f(x) = \left(\frac{1}{\alpha+1}\left(\alpha\gamma^{\frac{1}{\alpha+1}}x + (\alpha+1)\mathrm{e}^{\frac{\alpha\ \ln(\beta)}{\alpha+1}}\right)\right)^{\frac{\alpha+1}{\alpha}}.$$

A Class of Functions

- Since our use of W relies upon being able to solve

$$f'(x)^{\alpha+1} = \gamma f(x),$$

we ask for what kind of function f this is possible.

With initial condition $f(0) = \beta$, *Maple*'s built-in ODE solver returns

$$f(x) = \left(\frac{1}{\alpha+1}\left(\alpha\gamma^{\frac{1}{\alpha+1}}x + (\alpha+1)e^{\frac{\alpha\ \ln(\beta)}{\alpha+1}}\right)\right)^{\frac{\alpha+1}{\alpha}}.$$

- As $\alpha \to 0$, we retrieve the familiar $f(x) = \beta(\exp(\gamma x))$.
- Also, for $\alpha = 1, \gamma = 2$, when $\beta \to 0$, we recover $f(x) = \frac{x^2}{2}$.
- Thus, we obtain a large class of closed forms from which our previous examples emerge as limiting cases.

Simplified Closed Forms

- The closed form of the convex conjugates for functions of form $\beta \cdot \exp(x)$ simplifies to

$$g^*(y) = \begin{cases} y\left(\log(y) - W(y) - \frac{1}{W(y)} - \log(\beta)\right) & \text{if } y > 0 \\ -1 & \text{if } y = 0 \\ \infty & \text{if } y < 0 \end{cases}.$$

Simplified Closed Forms

- The closed form of the convex conjugates for functions of form $\beta \cdot \exp(x)$ simplifies to

$$g^*(y) = \begin{cases} y\left(\log(y) - W(y) - \frac{1}{W(y)} - \log(\beta)\right) & \text{if } y > 0 \\ -1 & \text{if } y = 0 \\ \infty & \text{if } y < 0 \end{cases}.$$

- Where $\frac{1}{q} + \frac{1}{p} = 1$, the closed form of the convex conjugates for functions of form $f(x) = \frac{|x|^p}{p}, (p > 1)$ simplifies to

$$g^*(y) = |y|\left(\left(\frac{p}{q}W\left(\frac{q}{p}|y|^q\right)\right)^{\frac{1}{p}} - \left(\frac{p}{q}W\left(\frac{q}{p}|y|^q\right)\right)^{-\frac{1}{q}}\right).$$

- Compare the former to the case $\beta = 1$ and the latter to the case $p = q = 2$, both of which we have seen before.

Conjugates of Compositions

Theorem (Conjugates of Compositions)

Consider the convex composition $h \circ g$ *of a nondecreasing convex function* $h : (-\infty, \infty] \to (-\infty, \infty]$ *with a convex function* $f : X \to (-\infty, \infty]$. *We interpret* $f(+\infty) = +\infty$, *and we assume there is a point* $\hat{x}$ *in* X *satisfying* $f(\hat{x}) \in \operatorname{int} \operatorname{dom}(h)$.
For y *in* X^*,

$$(h \circ f)^*(y) = \inf_{t \geq 0} \left\{ h^*(t) + t f^* \left(\frac{y}{t} \right) \right\}.$$

Here $0 f^* \left(\frac{y}{0} \right) = \iota^*_{\operatorname{dom} f}(y)$ *in terms of the convex indicator function* $\iota^*_{\operatorname{dom} f}$ *which is zero on* $\operatorname{dom} f$ *and is* $+\infty$ *otherwise.*

An Example with Composition

We may use Theorem 5 with

$$h(t) = \exp(t),\ h^*(t) = t \log t - t \text{ (the Shannon entropy)}$$

to compute the conjugate for $g(x) = \exp \circ f(x)$ for various f.

An Example with Composition

We may use Theorem 5 with

$$h(t) = \exp(t),\ h^*(t) = t \log t - t \text{ (the Shannon entropy)}$$

to compute the conjugate for $g(x) = \exp \circ f(x)$ for various f. For example, with $f(x) = \frac{|x|^p}{p}$, we may obtain g^* by evaluating $(h \circ f)^*$.

An Example with Composition

We may use Theorem 5 with

$$h(t) = \exp(t),\ h^*(t) = t \log t - t \text{ (the Shannon entropy)}$$

to compute the conjugate for $g(x) = \exp \circ f(x)$ for various f. For example, with $f(x) = \frac{|x|^p}{p}$, we may obtain g^* by evaluating $(h \circ f)^*$. From Theorem 5 we have that, for $y \neq 0$,

$$(h \circ f)^*(y) = \inf_{t \geq 0} \left\{ h^*(t) + t f^* \left(\frac{y}{t} \right) \right\} = \inf_{t \geq 0} \left\{ t \log t - t + t \left(\frac{|y|}{t} \right)^q / q \right\}.$$

An Example with Composition

We may use Theorem 5 with

$$h(t) = \exp(t),\ h^*(t) = t\log t - t \text{ (the Shannon entropy)}$$

to compute the conjugate for $g(x) = \exp \circ f(x)$ for various f. For example, with $f(x) = \frac{|x|^p}{p}$, we may obtain g^* by evaluating $(h \circ f)^*$. From Theorem 5 we have that, for $y \neq 0$,

$$(h \circ f)^*(y) = \inf_{t \geq 0} \left\{ h^*(t) + t f^* \left(\frac{y}{t} \right) \right\} = \inf_{t \geq 0} \left\{ t \log t - t + t \left(\frac{|y|}{t} \right)^q / q \right\}.$$

Differentiating, setting equal to zero, and solving for t, we arrive at

$$t = \exp\left(\frac{W\left((q-1)|y|^q\right)}{q} \right)$$

which we substitute to obtain the same answer as before.

Infimal Convolution

Consider for $\mu > 0$ the convolutions

$$g_\mu = (x \to x\log(x) - x)\,\square\,\mu\left(x \to \frac{x^2}{2}\right).$$

This family – of everywhere continuous functions – is also called the Moreau envelope of $x\log(x) - x$.

Infimal Convolution

Consider for $\mu > 0$ the convolutions

$$g_\mu = (x \to x\log(x) - x) \,\square\, \mu \left(x \to \frac{x^2}{2}\right).$$

This family – of everywhere continuous functions – is also called the Moreau envelope of $x\log(x) - x$.

SCAT provides:

$$g_\mu(y) = \frac{\mu}{2}y^2 - \frac{1}{\mu}W(\mu e^{\mu y}) - \frac{1}{2\mu}W(\mu e^{\mu y})^2.$$

Infimal Convolution

Consider for $\mu > 0$ the convolutions

$$g_\mu = (x \to x\log(x) - x) \,\square\, \mu \left(x \to \frac{x^2}{2}\right).$$

This family – of everywhere continuous functions – is also called the Moreau envelope of $x\log(x) - x$.

SCAT provides:

$$g_\mu(y) = \frac{\mu}{2}y^2 - \frac{1}{\mu}W(\mu e^{\mu y}) - \frac{1}{2\mu}W(\mu e^{\mu y})^2.$$

- g_μ is fully explicit in terms of W.

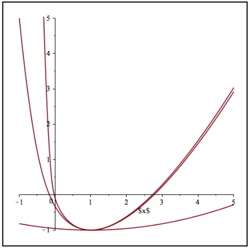


Figure: Convolution of entropy $x\log x - x$ and energy $\mu\, x^2/2$ for $\mu = 1/10, 10, 100$.

Homotopy

Consider for $0 \leq t \leq 1$ the combination

$$f_t(x) = (1-t)(x \log x - x) + t\frac{x^2}{2} \tag{13}$$

so that f_0 is the Shannon entropy and f_1 the energy.

Homotopy

Consider for $0 \leq t \leq 1$ the combination

$$f_t(x) = (1-t)(x \log x - x) + t\frac{x^2}{2} \tag{13}$$

so that f_0 is the Shannon entropy and f_1 the energy.

The conjugate of (13) is

$$f_t^*(y) = \frac{(1-t)^2}{2t}\left(W\left(\frac{t}{1-t}e^{\frac{y}{1-t}}\right)+2\right)W\left(\frac{t}{1-t}e^{\frac{y}{1-t}}\right).$$

In the limit at $t=1$ we recover the positive energy which is infinite for $y<0$ and at $t=0$ we reobtain $x\log(x)-x$.

Minimization with Constraints

Consider the (negative) entropy functional $I_f : L^1([0,1],\lambda) \to \mathbb{R}$ defined as follows:

$$I_f(x) = \int_0^1 f(x(s))\,\mathrm{d}s$$

where λ is Lebesgue measure and f is a proper, closed convex function.

Suppose we wish to minimize I_f subject to finitely many continuous linear constraints of the form

$$\langle a_k, x\rangle = \int_0^1 a_k(s)x(s)\,\mathrm{d}s = b_k$$

for $1 \leq k \leq n$. We may write this for $A : L^1([0,1]) \to \mathbb{R}^n$ with

$$Ax = \left(\int_0^1 a_1(s)x(s)\,\mathrm{d}s, \ldots, \int_0^1 a_n(s)x(s)\,\mathrm{d}s\right) = b.$$

Reformulation as Dual Problem

When f^* is smooth and everywhere finite on the real line, our problem

$$\inf_{x \in L^1} \{I_f(x) | Ax = b\}$$

reduces – via subtle Fenchel duality – to solving a finite nonlinear equation.

Solve for $\lambda_1, \lambda_2, \ldots, \lambda_n$

$$\int_0^1 \overbrace{(f^*)'\left(\sum_{j=1}^n \lambda_j a_j(s)\right)}^{\text{primal solution } x(s)} a_k(s)\,\mathrm{ds} = \mathrm{b_k} \qquad (1 \leq \mathrm{k} \leq \mathrm{n}). \tag{14}$$

Reformulation as Dual Problem

When f^* is smooth and everywhere finite on the real line, our problem

$$\inf_{x \in L^1} \{I_f(x) | Ax = b\}$$

reduces – via subtle Fenchel duality – to solving a finite nonlinear equation.

Solve for $\lambda_1, \lambda_2, \ldots, \lambda_n$

$$\int_0^1 \overbrace{(f^*)'\left(\sum_{j=1}^n \lambda_j a_j(s)\right)}^{\text{primal solution } x(s)} a_k(s)\,\mathrm{ds} = \mathrm{b_k} \qquad (1 \le \mathrm{k} \le \mathrm{n}). \tag{14}$$

Details are reprised in the paper accompanying this talk. More information – including the matter of primal attainment and constraint qualification – can be found in [9].

The Role of Lambert W

To illustrate the role of for W, we choose f in our optimization problem to be of the form

$$f_t(x) = (1-t)(x\log x - x) + t\frac{x^2}{2}.$$

The Role of Lambert W

To illustrate the role of for W, we choose f in our optimization problem to be of the form

$$f_t(x) = (1-t)(x \log x - x) + t\frac{x^2}{2}.$$

Then we have the following:

- f_0 is the Shannon Entropy
- f_1 is the energy
- $(f_t^*)'(y) = \dfrac{(1-t)}{t} W\left(\frac{t}{1-t} \exp\left(\dfrac{y}{1-t}\right)\right)$
- $\lim_{t\to 0} (f_t^*)'(y) = \exp(y)$
- $\lim_{t\to 1} (f_t^*)'(y) = \max\{y, 0\}$.

A Computational Example

We illustrate by implementing a program with m algebraic moments of the form

$$a_k(s) = s^{k-1} \qquad (k = 1 \ldots m).$$

A Computational Example

We illustrate by implementing a program with m algebraic moments of the form

$$a_k(s) = s^{k-1} \qquad (k = 1 \ldots m).$$

Our subgradient (dual problem) is represented more explicitly by following the set of equations for $k = 1 \ldots 10$:

$$\int_0^1 \frac{(1-t)}{t} W\left(\frac{t}{1-t} \exp\left(\frac{\sum_{j=1}^{n} \lambda_j s^{j-1}}{1-t}\right)\right) s^{k-1} \mathrm{d}s - b_k = 0. \tag{15}$$

We can solve for λ using any standard numerical solver or, say, by a Newton-type method.

Cost-Effective Computing

Newton's method is cost-effective for this formulation. The Hessian is a Hankel matrix:

$$H(\lambda) = (h_{i,k})$$

$$h_{i,k} = \int_0^1 \frac{(1-t)}{t} W\left(\frac{t}{1-t}\exp\left(\frac{\sum_{j=1}^n \lambda_j a_j(s)}{1-t}\right)\right) a_k(s)a_i(s)\mathrm{d}s$$

$$= \int_0^1 \frac{(1-t)}{t} W\left(\frac{t}{1-t}\exp\left(\frac{\sum_{j=1}^n \lambda_j s^{j-1}}{1-t}\right)\right) s^{k+i-2}\mathrm{d}s.$$

Cost-Effective Computing

Newton's method is cost-effective for this formulation. The Hessian is a Hankel matrix:

$$H(\lambda) = (h_{i,k})$$

$$h_{i,k} = \int_0^1 \frac{(1-t)}{t} W\left(\frac{t}{1-t}\exp\left(\frac{\sum_{j=1}^n \lambda_j a_j(s)}{1-t}\right)\right) a_k(s)a_i(s)\mathrm{d}s$$

$$= \int_0^1 \frac{(1-t)}{t} W\left(\frac{t}{1-t}\exp\left(\frac{\sum_{j=1}^n \lambda_j s^{j-1}}{1-t}\right)\right) s^{k+i-2}\mathrm{d}s.$$

- When m is the number of moments specified, for each iteration we need only to compute the $2m-1$ cases $k+i=2\ldots 2m$.
- The gradient $G(\lambda)$ may be obtained by taking the first row (or column) of the Hessian and subtracting b_k from the kth entry.

Saving Computation on the Quadrature

We adopt a Gaussian quadrature rule with weights $\{a_l\}_{l=1}^m$ and abcissas $\{x_l\}_{l=1}^m$. Then, where

$$F(x_l) = \frac{(1-t)}{t} W\left(\frac{t}{1-t}\exp\left(\frac{\sum_{j=1}^n \lambda_j x_l^{j-1}}{1-t}\right)\right),$$

for a single iteration we need only use numerical integration on the W function m times rather than order $m \cdot n$ times.

Saving Computation on the Quadrature

We adopt a Gaussian quadrature rule with weights $\{a_l\}_{l=1}^m$ and abcissas $\{x_l\}_{l=1}^m$. Then, where

$$F(x_l) = \frac{(1-t)}{t} W\left(\frac{t}{1-t}\exp\left(\frac{\sum_{j=1}^n \lambda_j x_l^{j-1}}{1-t}\right)\right),$$

for a single iteration we need only use numerical integration on the W function m times rather than order $m \cdot n$ times.

- To see more clearly why this is the case, notice that we can reuse the values $a_l F(x_l)$, $l = 1 \ldots m$ as follows:

$$h_{1,1} = \sum_{l=0}^m a_l F(x_l), h_{(i+k=\alpha)} = \sum_{l=0}^m a_l F(x_l) x_l^{\alpha-2}.$$

- Thus, we need only compute each once for each iteration. We can also reuse $x_l^{\alpha-2}$ for $l = 1 \ldots m$, $\alpha = 2 \ldots 20$.

Complete Optimized Process

Our full method for computing with minimal cost is as follows:

Complete Optimized Process

Our full method for computing with minimal cost is as follows:

1. Precompute the weights $\{a_l\}_{l=1}^{m}$, and the abscissas raised to various powers x_l^{α}, $l = 1 \ldots m, \alpha = 0 \ldots 18$, storing the weights in a vector and the powers of the abscissas in a matrix.

Complete Optimized Process

Our full method for computing with minimal cost is as follows:

1. Precompute the weights $\{a_l\}_{l=1}^{m}$, and the abscissas raised to various powers x_l^{α}, $l = 1 \ldots m, \alpha = 0 \ldots 18$, storing the weights in a vector and the powers of the abscissas in a matrix.
2. At each step compute the function values $a_l F(x_l)$, $l = 1 \ldots m$, storing them in a vector.

Complete Optimized Process

Our full method for computing with minimal cost is as follows:

1. Precompute the weights $\{a_l\}_{l=1}^m$, and the abscissas raised to various powers x_l^α, $l = 1 \ldots m, \alpha = 0 \ldots 18$, storing the weights in a vector and the powers of the abscissas in a matrix.
2. At each step compute the function values $a_l F(x_l)$, $l = 1 \ldots m$, storing them in a vector.
3. Compute the necessary 19 Hessian values $\sum_{l=0}^m a_l F(x_l) x_l^{\alpha-2}$, $\alpha = 2 \ldots 20$. If we properly create our matrix – of stored abscissa values raised to powers – we will be able to compute the Hessian values by simply multiplying our vector from Step 2 by this matrix.

Complete Optimized Process

Our full method for computing with minimal cost is as follows:

1. Precompute the weights $\{a_l\}_{l=1}^{m}$, and the abscissas raised to various powers x_l^α, $l = 1 \dots m, \alpha = 0 \dots 18$, storing the weights in a vector and the powers of the abscissas in a matrix.
2. At each step compute the function values $a_l F(x_l)$, $l = 1 \dots m$, storing them in a vector.
3. Compute the necessary 19 Hessian values $\sum_{l=0}^{m} a_l F(x_l) x_l^{\alpha-2}$, $\alpha = 2 \dots 20$. If we properly create our matrix – of stored abscissa values raised to powers – we will be able to compute the Hessian values by simply multiplying our vector from Step 2 by this matrix.
4. Use the resultant 19 values to build the Hessian and gradient and then solve for the next iterate.

For consistency, all examples in this subsection used:

- 24 digits of precision
- 20 abscissas
- A Newton step size of $1/2$
- 8 moments unless otherwise specified
- A t value of $\frac{1}{2}$ unless otherwise specified
- The objective function of $s \to \frac{6}{10} + sin(3\pi s^2)$ unless otherwise specified.

For consistency, all examples in this subsection used:

- 24 digits of precision
- 20 abscissas
- A Newton step size of $1/2$
- 8 moments unless otherwise specified
- A t value of $\frac{1}{2}$ unless otherwise specified
- The objective function of $s \to \frac{6}{10} + sin(3\pi s^2)$ unless otherwise specified.

This reduced step dramatically improved convergence for t near 1. While this precision is higher than would be used in production code, it allows us to see the optimal performance of the algorithm.

Visualizing Accuracy

We ask *Maple* to compute until the error, as measured by the norm of the gradient, is less than 10^{-10}. At 46 iterations we obtain λ values:

$$-0.7079161355,$$
$$10.64405426,$$
$$-126.5979784,$$
$$656.6020449,$$
$$-1458.868219,$$
$$1329.347874$$
$$-299.1180785,$$
$$-112.3114246$$

where the error is about $6.84330e-11$.

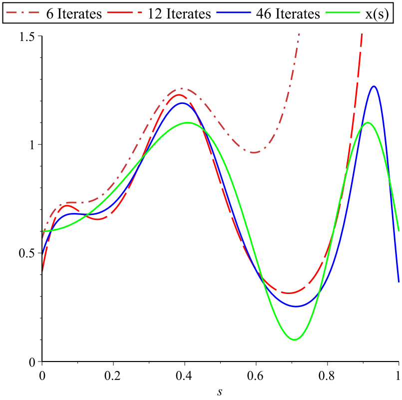


Figure: The primal solutions for iterates 6, 12, and 46.

Variation of t

- We consider five different possible values for t: $0, .25, .5, .75, 1$.
- We run Newton's Method for each case until meeting the requirement that the norm of the gradient is less than or equal to 10^{-10}.
- Notice that as t increases the visual fit increases substantially. One cannot determine this from looking at the numerical error alone.

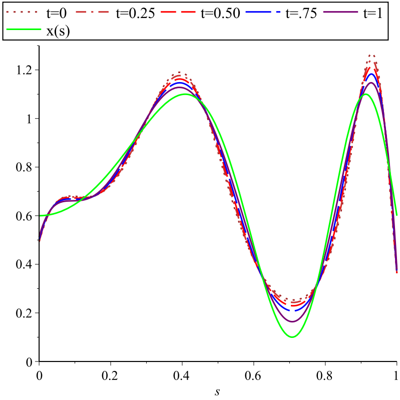
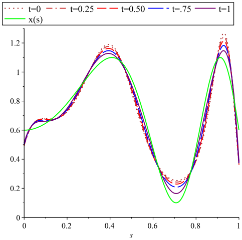


Figure: The associated primal solutions for various choices of t.

Solutions for Various Choices of t

t	0	.25	.5	.75	1
λ_1	-.707916	-.404828	-.101065	.204002	.512307
λ_2	10.6440	9.46383	8.23003	6.90162	5.36009
λ_3	-126.597	-114.651	-101.923	-87.8556	-70.8919
λ_4	656.602	605.686	550.755	488.934	412.561
λ_5	-1458.86	-1368.32	-1269.02	-1154.26	-1007.13
λ_6	1329.34	1282.68	1227.95	1157.70	1054.85
λ_7	-299.118	-329.937	-358.596	-381.447	-391.764
λ_8	-112.311	-85.1887	-57.6202	-30.1516	-3.12491
Error	6.84330e-11	9.81661e-11	8.26865e-11	9.6666e-11	7.05698e-11
Iterates	46	46	47	47	47

Dual solutions corresponding to various choices of t are shown in the Table while primal solutions are shown to the right.



Varying the Number of Moments

- We consider the choice of 4, 8, 12, and 20 moments.
- We run Newton's Method for each case until meeting the requirement that the norm of the gradient is less than or equal to 10^{-10}.
- While we used 26 digits of precision for all of these examples (for consistency), this was the only case wherein we used 20 moments and so exploited the employment of such high precision.

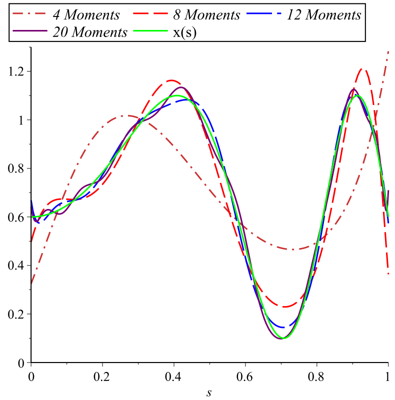


Figure: The primal solutions for various numbers of moments

Changing the Objective Function: A Pulse

- We compute with the pulse: $x(s) = \chi_{[0,\frac{1}{2}]}(s)$.
- The pulse is a more computationally challenging example because of its jump discontinuity and constancy on an open interval.
- This slowed the convergence of the gradient to zero with more moments, especially for values of t nearer to 1.
- The desired properties can still be seen visually.

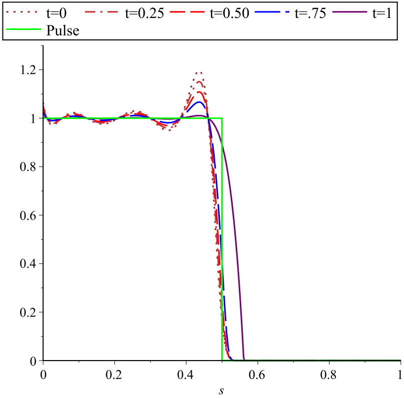
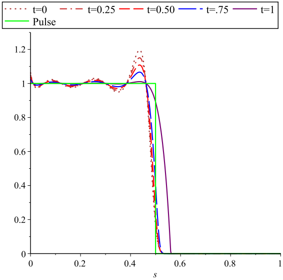


Figure: The primal solutions for various numbers of moments

Changing the Objective Function: A Pulse

t	0	.25	.5	.75	1
Error	6.87225e-11	7.45516e-11	9.69259e-11	1.9136e-11	.21252e-5
Iterates	70	62	55	48	200

- We instruct *Maple* to stop computing once the norm of the gradient is less than 10^{-10} or after reaching 200 iterates.
- For $t = 1$, we reached 200 iterates before the norm of the gradient was less than 10^{-10}, but the primal solution we obtained is still a good proxy for the pulse. This can be seen in the Figure, where the Gibbs Phenomenon may also be clearly observed for the the other values of t.



When a Closed Form is not Forthcoming

Even when one is not able to produce a closed form, *SCAT* and its numerical partner *CCAT* may still help.

Example: $f := (0, \infty) \to \mathbb{R}$ by

$$f(x) = \left(\frac{x}{e}\right)^x$$

SCAT does not return a closed form but still produces the plot shown.

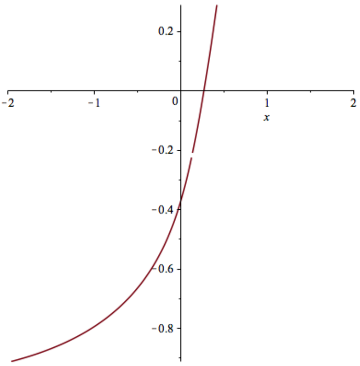


Figure: The Conjugate of f

Another Example: $\log \Gamma$

For the conjugate, SCAT returns:

$$RootOf\left(-\Psi\left(\_Z\right)+x\right)x-\log\left(\Gamma\left(RootOf\left(-\Psi\left(\_Z\right)+x\right)\right)\right).$$

where Ψ is the Psi function.

Maple's root finder struggles, leaving the plot incomplete. This can be obviated by a Newton solver for $x>0$ of $\Psi(x)=y$. Set

$$x_0=\begin{cases}\exp(y)+1/2 & \text{if } y\geq -2.2\\ -1/(y-\Psi(1)) & \text{otherwise}\end{cases}$$

$$x_{n+1}=x_n-\frac{\Psi(x_n)-y}{\Psi'(x_n)}.$$

– Ψ and Ψ' are also known as digamma and trigamma functions.

Another Example: $\log \Gamma$

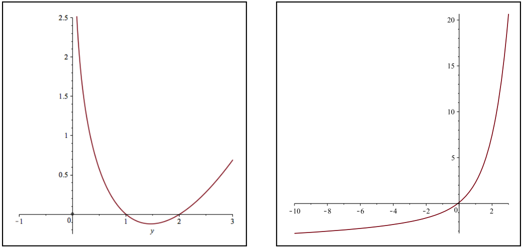


Figure: The function $\log \Gamma$ (L) and its conjugate (R).

- We hope that we have made a good advertisement for the value of W in optimisation and elsewhere.
- We also hope we have highlighted the usefulness of SCAT and its numerical partner CCAT.

References I

[1] J. Barzilai and J.M. Borwein, "Two point step-size methods," *IMA Journal on Numerical Analysis*, **8** (1988), 141–148.

[2] D. Bersetkas, "Projected Newton Methods for optimization problems with simple constraints, " *SIAM. J. Control and Optimization* **20**(12) (1988), 221–246.

[3] J.M. Borwein, "The SIAM 100 Digits Challenge," Extended review for the *Mathematical Intelligencer*, **27** (4) (2005), 40–48.

[4] J.M. Borwein and D.H. Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century* A.K. Peters Ltd, 2004. ISBN: 1-56881-136-5. Combined Interactive CD version 2006. Expanded Second Edition, 2008.

[5] J.M. Borwein, D.H. Bailey and R. Girgensohn, *Experimentation in Mathematics: Computational Paths to Discovery,* A.K. Peters Ltd, 2004. ISBN: 1-56881-211-6. Combined Interactive CD version 2006.

[6] J.M. Borwein and R.E. Crandall, "Closed forms: what they are and why we care." *Notices Amer. Math. Soc.* **60**:1 (2013), 50–65.

[7] J.M. Borwein and C. Hamilton, "Symbolic Convex Analysis: Algorithms and Examples," *Mathematical Programming*, **116** (2009), 17–35.

[8] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization : Theory and Examples*. Springer, (2000) (2nd Edition, 2006).

[9] J. M. Borwein and A. S. Lewis, "Duality relationships for entropy–like minimization problems." *SIAM Control and Optim.*, **29** (1991), 325–338.

References II

[10] Jonathan M. Borwein, Armin Straub, James Wan and Wadim Zudilin, with an Appendix by Don Zagier, "Densities of short uniform random walks." *Canadian. J. Math.* **64** (5), (2012), 961–990. Available at http://arxiv.org/abs/1103.2995.

[11] J.M. Borwein and J. D. Vanderwerff, *Convex Functions : Constructions, Characterizations and Counterexamples*. Cambridge University Press, (2010).

[12] Jonathan M. Borwein, Alf van der Poorten, Jeff Shallit, and Wadim Zudilin. *Neverending Fractions*, Australia Mathematical Society Lecture Series, Cambridge University Press. 2014.

[13] J.M. Borwein and Liangjin Yao, "Legendre-Type Integrands and Convex Integral Functions," *Journal of Convex Analysis*, **21** (2014), 264-288.

[14] J.M. Borwein and Qiji Zhu, *Techniques of Variational Analysis,* CMS/Springer-Verlag, 2005. Paperback, 2010.

[15] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*, Cambridge University Press. 2004.

[16] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth, "On the Lambert W Function," *Advances in Computational Mathematics*, **5** (1996), 329-359.

[17] Thomas P. Dence, "A Brief Look into the Lambert W Function," *Applied Mathematics*, **4** (2013), 887-892.

[18] D.E. Knuth, C.C. Rousseau, "A Stirling series: 10832," *American Mathematical Monthly*, **108** (2001), 877–878. http://dx.doi.org/10.2307/2695574.

[19] Scott Lindstrom, MSc Thesis. Available at https://docserver.carma.newcastle.edu.au/.

[20] Mila Mršević. "Convexity of the Inverse Function," *The Teaching of Mathematics*, **XI** (2008), 21-24.

[21] B.S. Mordukhovich and N.M. Nam, *An Easy Path to Convex Analysis*, Morgan & Claypool, 2014. ISBN: 9781627052375.

References III

[22] Stephen M. Stigler, "Stigler's law of eponymy". (F. Gieryn, ed.) "Stigler's law of eponymy." *Transactions of the New York Academy of Sciences* **39** (1980), 147?58. doi:10.1111/j.2164-0947.1980.tb02775.x.

[23] Croatia Encyclopedia: http://www.enciklopedija.hr/natuknica.aspx?ID=35243