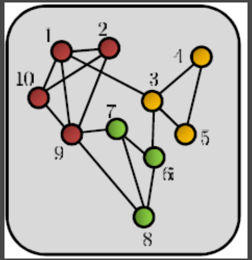

INFERENCE OVER NETWORKS

LECTURE #3: Complex Hessian Matrices

Professor Ali H. Sayed
UCLA Electrical Engineering

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Reference

Appendix B (Complex Hessian Matrices, pp. 720-729):

A. H. Sayed, ``Adaptation, learning, and optimization over networks," ***Foundations and Trends in Machine Learning***, vol. 7, issue 4-5, pp. 311-801, NOW Publishers, 2014.

Setting

Hessian matrices involve second-order partial derivatives, which we shall assume to be continuous functions of their arguments whenever necessary. Some effort is needed to define Hessian matrices for functions of complex variables. For this reason, we consider first the case of real arguments to help motivate the extension to complex arguments. In this appendix we only consider *real-valued* functions $g(z) \in \mathbb{R}$, which corresponds to the situation of most interest to us since utility or cost functions in adaptation and learning are generally real-valued.

Real Arguments

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Real Arguments

We continue to denote the individual entries of the column vector $z \in \mathbb{R}^M$ by $z = \text{col}\{z_1, z_2, \ldots, z_M\}$. The Hessian matrix of $g(z) \in \mathbb{R}$ is an $M \times M$ *symmetric* matrix function of z, denoted by $H(z)$, and whose $(m,n)-$th entry is constructed as follows:

$$[H(z)]_{m,n} \triangleq \frac{\partial^2 g(z)}{\partial z_m \partial z_n} = \frac{\partial}{\partial z_m}\left[\frac{\partial g(z)}{\partial z_n}\right] = \frac{\partial}{\partial z_n}\left[\frac{\partial g(z)}{\partial z_m}\right] \qquad \text{(B.1)}$$

in terms of the partial derivatives of $g(z)$ with respect to the real scalar arguments $\{z_m, z_n\}$.

Real Arguments

For example, for a two-dimensional argument z (i.e., $M = 2$), the four entries of the 2×2 Hessian matrix would be given by:

$$H(z) = \begin{bmatrix} \dfrac{\partial^2 g(z)}{\partial z_1^2} & \dfrac{\partial^2 g(z)}{\partial z_1 \partial z_2} \\[2ex] \dfrac{\partial^2 g(z)}{\partial z_2 \partial z_1} & \dfrac{\partial^2 g(z)}{\partial z_2^2} \end{bmatrix} \tag{B.2}$$

Real Arguments

It is straightforward to recognize that the Hessian matrix $H(z)$ defined by (B.1) can be obtained as the result of two successive gradient vector calculations with respect to z and z^{T} in the following manner (where the order of the differentiation does not matter):

$$H(z) \triangleq \nabla_{z^{\mathsf{T}}}\left[\nabla_z\, g(z)\right] = \nabla_z\left[\nabla_{z^{\mathsf{T}}}\, g(z)\right] \qquad (M \times M) \tag{B.3}$$

Real Arguments

For instance, using the first expression, the gradient operation $\nabla_z\, g(z)$ generates a $1 \times M$ (row) vector function and the subsequent differentiation with respect to z^{T} leads to the $M \times M$ Hessian matrix, $H(z)$. It is clear from (B.3) that the Hessian matrix is indeed symmetric so that

$$H(z) \;=\; H^{\mathsf{T}}(z) \tag{B.4}$$

Real Arguments

A useful property of Hessian matrices is that they help characterize the nature of stationary points of functions $g(z)$ that are twice continuously differentiable. Specifically, if z^o is a stationary point of $g(z)$ (i.e., a point where $\nabla_z\, g(z) = 0$), then the following facts hold (see, e.g., [36, 93]):

(a) z^o is a local minimum of $g(z)$ if $H(z^o) > 0$, i.e., if all eigenvalues of $H(z^o)$ are positive.

(b) z^o is a local maximum of $g(z)$ if $H(z^o) < 0$, i.e., if all eigenvalues of $H(z^o)$ are negative.

Example #B.1

Example B.1 (Quadratic cost functions). Consider the quadratic function

$$g(z) = \kappa + a^{\mathsf{T}} z + z^{\mathsf{T}} b + z^{\mathsf{T}} C z \tag{B.5}$$

where κ is a scalar, $\{a, b\}$ are column vectors of dimension $M \times 1$ each, and C is an $M \times M$ *symmetric* matrix (all of them are real-valued in this case).

Example #B.1

We know from (A.22) and (A.30) that any stationary point, z^o, of $g(z)$ should satisfy the linear system of equations

$$Cz^o = \frac{1}{2}(a+b) \tag{B.6}$$

It follows that z^o is unique if, and only if, C is nonsingular. Moreover, in this case, the Hessian matrix is given by

$$H = 2C \tag{B.7}$$

which is independent of z. It follows that the quadratic function $g(z)$ will have a *unique* global minimum if, and only if, $C > 0$.

■

Complex Arguments

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Complex Arguments

We now extend the definition of Hessian matrices to functions $g(z) \in \mathbb{R}$ that are still *real-valued* but their argument, $z \in \mathbb{C}^M$, is complex-valued. This case is of great interest in adaptation, learning, and estimation problems since cost functions are generally real-valued while their arguments can be complex-valued. The Hessian matrix of $g(z)$ can now be defined in two equivalent forms by working either with the complex variables $\{z, z^*\}$ directly or with the real and imaginary parts $\{x, y\}$ of z. In contrast to the case of real arguments studied above in (B.3), where the Hessian matrix had dimensions $M \times M$, the Hessian matrix for complex arguments will be twice as large.

Complex Arguments

We start by expressing each entry z_m of z in terms of its real and imaginary components as

$$z_m = x_m + jy_m, \quad m = 1, 2, \ldots, M \tag{B.8}$$

We subsequently collect the real and imaginary factors $\{x_m\}$ and $\{y_m\}$ into two real vectors:

$$x \overset{\Delta}{=} \text{col}\{x_1, x_2, \ldots, x_M\} \tag{B.9}$$

$$y \overset{\Delta}{=} \text{col}\{y_1, y_2, \ldots, y_M\} \tag{B.10}$$

Complex Arguments

so that

$$z = x + jy \qquad \text{(B.11)}$$

Then, we can equivalently express $g(z)$ as a function of $2M$ *real* variables as $g(z) = g(x, y)$. We now proceed to define the Hessian matrix of $g(z)$ in two equivalent ways by working with either the complex variables $\{z, z^*\}$ or the real variables $\{x, y\}$. We consider the latter case first since we can then call upon the earlier definition (B.3) for real arguments.

Real Hessian Matrix

Since $g(x, y) \in \mathbb{R}$ is a function of the real arguments $\{x, y\}$, we can invoke definition (B.3) to associate with $g(x, y)$ a *real* Hessian matrix $H(x, y)$; its dimensions will be $2M \times 2M$. This Hessian matrix will involve second-order partial derivatives relative to x and y. For example, when $z = x + jy$ is a *scalar*, then $H(x, y)$ will be 2×2 and given by:

$$H(x,y) \;=\; \begin{bmatrix} \dfrac{\partial^2 g(x,y)}{\partial x^2} & \dfrac{\partial^2 g(x,y)}{\partial x \partial y} \\[2ex] \dfrac{\partial^2 g(x,y)}{\partial y \partial x} & \dfrac{\partial^2 g(x,y)}{\partial y^2} \end{bmatrix}, \quad z = x + jy \qquad \text{(B.12)}$$

Real Hessian Matrix

Likewise, when z is two-dimensional (i.e., $M = 2$) with entries $z_1 = x_1 + jy_1$ and $z_2 = x_2 + jy_2$, then the Hessian matrix of $g(z)$ will be 4×4 and given by:

Real Hessian Matrix

$$H(x,y) = \left[\begin{array}{cc|cc} \frac{\partial^2 g(z)}{\partial x_1^2} & \frac{\partial^2 g(z)}{\partial x_1 \partial x_2} & \frac{\partial^2 g(z)}{\partial x_1 \partial y_1} & \frac{\partial^2 g(z)}{\partial x_1 \partial y_2} \\ \frac{\partial^2 g(z)}{\partial x_2 \partial x_1} & \frac{\partial^2 g(z)}{\partial x_2^2} & \frac{\partial^2 g(z)}{\partial x_2 \partial y_1} & \frac{\partial^2 g(z)}{\partial x_2 \partial y_2} \\ \hline \frac{\partial^2 g(z)}{\partial y_1 \partial x_1} & \frac{\partial^2 g(z)}{\partial y_1 \partial x_2} & \frac{\partial^2 g(z)}{\partial y_1^2} & \frac{\partial^2 g(z)}{\partial y_1 \partial y_2} \\ \frac{\partial^2 g(z)}{\partial y_2 \partial x_1} & \frac{\partial^2 g(z)}{\partial y_2 \partial x_2} & \frac{\partial^2 g(z)}{\partial y_2 \partial y_1} & \frac{\partial^2 g(z)}{\partial y_2^2} \end{array}\right] \qquad \text{(B.13)}$$

Real Hessian Matrix

More generally, for arguments $z = x + jy$ of arbitrary dimensions $M \times 1$, the real Hessian matrix of $g(z)$ can be expressed in partitioned form in terms of 4 sub-matrices of size $M \times M$ each:

Real Hessian Matrix

$$
H(x,y) = \left[\begin{array}{c|c} \nabla_{x^{\mathsf T}}[\nabla_x \, g(x,y)] & \nabla_{x^{\mathsf T}}[\nabla_y \, g(x,y)] \\ \hline \nabla_{y^{\mathsf T}}[\nabla_x \, g(x,y)] & \nabla_{y^{\mathsf T}}[\nabla_y \, g(x,y)] \end{array}\right]
$$

$$
\triangleq \left[\begin{array}{c|c} H_{x^{\mathsf T}x} & \left(H_{y^{\mathsf T}x}\right)^{\mathsf T} \\ \hline H_{y^{\mathsf T}x} & H_{y^{\mathsf T}y} \end{array}\right] \qquad \text{(B.14)}
$$

Real Hessian Matrix

where we introduced the compact notation $\{H_{x^\mathsf{T}x}, H_{y^\mathsf{T}y}, H_{y^\mathsf{T}x}\}$ to denote the following second-order differentiation operations relative to the variables x and y:

$$\begin{cases} H_{x^\mathsf{T}x} & \triangleq & \nabla_{x^\mathsf{T}}[\nabla_x \, g(x,y)] \\ H_{y^\mathsf{T}y} & \triangleq & \nabla_{y^\mathsf{T}}[\nabla_y \, g(x,y)] \\ H_{y^\mathsf{T}x} & \triangleq & \nabla_{y^\mathsf{T}}[\nabla_x \, g(x,y)] \end{cases} \tag{B.15}$$

Real Hessian Matrix

We can express result (B.14) more compactly by working with the $2M \times 1$ extended vector v that is obtained by stacking x and y into a single vector:

$$v \triangleq \text{col}\{x, y\} \tag{B.16}$$

Then, the function $g(z)$ can also be regarded as a function of v, namely, $g(v)$. It is straightforward to verify that the same Hessian matrix $H(x, y)$ given by (B.14) can be expressed in terms of differentiation of $g(v)$ with respect to v as follows (compare with (B.3)):

Real Hessian Matrix

$$H(v) \stackrel{\Delta}{=} \nabla_{v^{\mathsf{T}}}\left[\nabla_v\, g(v)\right] = \nabla_v\left[\nabla_{v^{\mathsf{T}}}\, g(v)\right] \;=\; H(x,y) \qquad (2M \times 2M) \tag{B.17}$$

We shall use the alternative representation $H(v)$ more frequently than $H(x,y)$ and refer to it as the *real* Hessian matrix. It is clear from expressions (B.14) or (B.17) that the Hessian matrix so defined is symmetric so that

$$H(v) \;=\; H^{\mathsf{T}}(v) \tag{B.18}$$

Real Hessian Matrix

Again, a useful property of the Hessian matrix is that it can be used to characterize the nature of stationary points of functions $g(z)$ that are twice continuously differentiable. Specifically, if $z^o = x^o + jy^o$ is a stationary point of $g(z)$ (i.e., a point where $\nabla_z\, g(z) = 0$), then the following facts hold for $v^o = \text{col}\{x^o, y^o\}$:

(a) z^o is a local minimum of $g(z)$ if $H(v^o) > 0$, i.e., all eigenvalues of $H(v^o)$ are positive.

(b) z^o is a local maximum of $g(z)$ if $H(v^o) < 0$, i.e., all eigenvalues of $H(v^o)$ are negative.

Complex Hessian Matrix

Besides $H(v)$, we can associate a second Hessian matrix representation with $g(z)$ by working directly with the complex variables z and z^* rather than their real and imaginary parts, x and y (or v). We refer to this second representation as the *complex* Hessian matrix and we denote it by $H_c(z)$, with the subscript "c" used to distinguish it from the real Hessian matrix, $H(v)$, defined by (B.17). The complex Hessian, $H_c(z)$, is still $2M \times 2M$ and its four block partitions are now defined in terms of (Wirtinger) complex gradient operations relative to the variables z and z^* as follows (compare with (B.14)):

Complex Hessian Matrix

$$H_c(z) \triangleq \left[\begin{array}{c|c} H_{z^*z} & (H_{z^\mathsf{T}z})^* \\ \hline H_{z^\mathsf{T}z} & (H_{z^*z})^\mathsf{T} \end{array}\right] \quad (2M \times 2M) \qquad \text{(B.19)}$$

where the $M \times M$ block matrices $\{H_{z^*z}, H_{z^\mathsf{T}z}\}$ correspond to the operations:

$$\begin{cases} H_{z^*z} \triangleq \nabla_{z^*}[\nabla_z\, g(z)] \\ H_{z^\mathsf{T}z} \triangleq \nabla_{z^\mathsf{T}}[\nabla_z\, g(z)] \end{cases} \qquad \text{(B.20)}$$

Complex Hessian Matrix

It is clear from definition (B.19) that the complex Hessian matrix is now Hermitian so that

$$H_c(z) \;=\; [H_c(z)]^* \tag{B.21}$$

For example, for the same case (B.12) when z is a scalar, definition (B.19) leads to:

$$H_c(z) \;=\; \begin{bmatrix} \dfrac{\partial^2 g(z)}{\partial z^* \partial z} & \dfrac{\partial^2 g(z)}{\partial z^{*2}} \\[2ex] \dfrac{\partial^2 g(z)}{\partial z^2} & \dfrac{\partial^2 g(z)}{\partial z \partial z^*} \end{bmatrix} \tag{B.22}$$

Complex Hessian Matrix

Likewise, for the two-dimensional case (B.13), the complex Hessian matrix is given by:

$$H_c(z) = \left[\begin{array}{cc|cc} \dfrac{\partial^2 g(z)}{\partial z_1^* \partial z_1} & \dfrac{\partial^2 g(z)}{\partial z_1^* \partial z_2} & \dfrac{\partial^2 g(z)}{\partial z_1^{*2}} & \dfrac{\partial^2 g(z)}{\partial z_1^* \partial z_2^*} \\ \dfrac{\partial^2 g(z)}{\partial z_2^* \partial z_1} & \dfrac{\partial^2 g(z)}{\partial z_2^* \partial z_2} & \dfrac{\partial^2 g(z)}{\partial z_2^* \partial z_1^*} & \dfrac{\partial^2 g(z)}{\partial z_2^{*2}} \\ \hline \dfrac{\partial^2 g(z)}{\partial z_1^2} & \dfrac{\partial^2 g(z)}{\partial z_1 \partial z_2} & \dfrac{\partial^2 g(z)}{\partial z_1 \partial z_1^*} & \dfrac{\partial^2 g(z)}{\partial z_1 \partial z_2^*} \\ \dfrac{\partial^2 g(z)}{\partial z_2 \partial z_1} & \dfrac{\partial^2 g(z)}{\partial z_2^2} & \dfrac{\partial^2 g(z)}{\partial z_2 \partial z_1^*} & \dfrac{\partial^2 g(z)}{\partial z_2 \partial z_2^*} \end{array}\right] \qquad \text{(B.23)}$$

Complex Hessian Matrix

Observe further that if we introduce the $2M \times 1$ extended vector:

$$u \stackrel{\Delta}{=} \text{col}\left\{ z, (z^*)^{\mathsf{T}} \right\} \tag{B.24}$$

then we can express $H_c(z)$ in the following equivalent form in terms of the variable u (compare with (B.17)):

$$H_c(u) \stackrel{\Delta}{=} \nabla_{u^*}\left[\nabla_u \, g(u)\right] = \nabla_u\left[\nabla_{u^*} \, g(u)\right] = H_c(z) \quad (2M \times 2M) \tag{B.25}$$

Relating Both Representations

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Relation between Representations

The two Hessian forms, $H(v)$ and $H_c(u)$, defined by (B.17) and (B.25) are closely related to each other. Indeed, using (A.10), it can be verified that [219, 252]:

$$\begin{cases} H_c(u) &= \frac{1}{4} D H(v) D^* \\ H(v) &= D^* H_c(u) D \end{cases} \tag{B.26}$$

Relation between Representations

where D is the following $2M \times 2M$ block matrix:

$$D \triangleq \begin{bmatrix} I_M & jI_M \\ I_M & -jI_M \end{bmatrix} \tag{B.27}$$

where I_M denotes the identity matrix of size M. It is straightforward to verify that

$$DD^* = 2\, I_{2M} \tag{B.28}$$

so that D is almost unitary (apart from scaling by $1/\sqrt{2}$).

Relation between Representations

It follows from (B.26) and (B.28) that the matrices $H_c(u)$ and $\frac{1}{2}H(v)$ are similar to each other and, hence, the eigenvalues of $H_c(u)$ coincide with the eigenvalues of $\frac{1}{2}H(v)$ [104, 113]. We conclude that the *complex* Hessian matrix, $H_c(u)$, can also be used to characterize the nature of stationary points of $g(z)$, just like it was the case with the *real* Hessian matrix, $H(v)$. Specifically, if $z^o = x^o + jy^o$ is a stationary point of $g(z)$ (i.e., a point where $\nabla_z\, g(z) = 0$), then the following facts hold:

Relation between Representations

(a) z^o is a local minimum of $g(z)$ if $H_c(u^o) > 0$, i.e., all eigenvalues of $H_c(u^o)$ are positive.

(b) z^o is a local maximum of $g(z)$ if $H_c(u^o) < 0$, i.e., all eigenvalues of $H_c(u^o)$ are negative.

where $u^o = \text{col}\left\{z^o, (z^{o*})^{\mathsf{T}}\right\}$.

Relation between Representations

For ease of reference, Table B.1 summarizes the various definitions of Hessian matrices for *real-valued* functions $g(z) \in \mathbb{R}$ for both cases when z is real or complex-valued. In the latter case, there are two equivalent representations for the Hessian matrix: one representation is in terms of the real components $\{x, y\}$ and the second representation is in terms of the complex components $\{z, z^*\}$. The Hessian matrix has dimensions $M \times M$ when z is real and $2M \times 2M$ when z is complex. It is customary to use the compact notation $\nabla_z^2\, g(z)$ to refer to the Hessian matrix whether z is real or complex and by that notation we mean the following:

Summary

Table B.1: Definition of Hessian matrices for real-valued functions $g(z) \in \mathbb{R}$ for both cases when z is real-valued or complex-valued.

	Hessian matrix	**variables**	**dimensions**
z **real**	$H(z) = \nabla_{z^{\mathsf{T}}}[\nabla_z \, g(z)]$		$M \times M$
z **complex**	$H(v) = \nabla_{v^{\mathsf{T}}}[\nabla_v \, g(v)]$	$v = \begin{bmatrix} x \\ y \end{bmatrix}$	$2M \times 2M$
$z = x + jy$	$H_c(u) = \nabla_{u^*}[\nabla_u \, g(u)]$	$u = \begin{bmatrix} z \\ (z^*)^{\mathsf{T}} \end{bmatrix}$	

Summary

$$\nabla_z^2\, g(z) \;\stackrel{\Delta}{=}\; \begin{cases} \nabla_{z^{\mathsf{T}}}[\nabla_z\, g(z)], & \text{when } z \text{ is real} \;\; (M \times M) \\ \nabla_{u^*}[\nabla_u\, g(u)], & \text{when } z \text{ is complex} \;\; (2M \times 2M) \end{cases} \quad \text{(B.29)}$$

Example #B.2

Example B.2 (Hessian matrix calculations). Let us illustrate the above definitions by considering a couple of examples.

(1) Let $g(z) = |z|^2 = x^2 + y^2$, where z is a scalar. Then,

$$H(v) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \equiv H, \qquad H_c(u) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \equiv H_c \tag{B.30}$$

In this case, the Hessian matrices turn out to be constant and independent of v and u.

Example #B.2

(2) Consider now

$$g(z) = |z_1|^2 + 2\,\mathrm{Re}(z_1^* z_2) \;=\; x_1^2 + y_1^2 + 2x_1x_2 + 2y_1y_2 \qquad \text{(B.31)}$$

where $z = \mathrm{col}\{z_1, z_2\}$ is 2×1. Then, the Hessian matrices are again independent of v and u:

$$H(v) = \begin{bmatrix} 2 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix} \equiv H, \qquad H_c(u) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \equiv H_c \quad \text{(B.32)}$$

Example #B.2

(3) Consider the real-valued quadratic function:

$$g(z) = \kappa + a^* z + z^* a + z^* C z \tag{B.33}$$

where κ is a real scalar, a is a column vector, and C is a Hermitian matrix. Then,

$$H_{z^* z} = \nabla_{z^*}\left[\nabla_z\, g(z)\right] = C \tag{B.34}$$

$$H_{z^\mathsf{T} z} = \nabla_{z^\mathsf{T}}\left[\nabla_z\, g(z)\right] = 0 \tag{B.35}$$

Example #B.2

so that

$$H_c(u) \;=\; \begin{bmatrix} C & 0 \\ 0 & C^{\mathsf{T}} \end{bmatrix} \equiv H_c \qquad \text{(B.36)}$$

$$H(v) \;=\; \begin{bmatrix} C + C^{\mathsf{T}} & j(C - C^{\mathsf{T}}) \\ j(C^{\mathsf{T}} - C) & C + C^{\mathsf{T}} \end{bmatrix} \equiv H \qquad \text{(B.37)}$$

It follows from the expression for $H_c(u)$ that it is sufficient to examine the inertia of C to determine the nature of the stationary point(s) of $g(z)$.

■

Example #B.3

Example B.3 (Block diagonal Hessian matrix). Observe from definition (B.19) that the complex Hessian matrix becomes *block diagonal* whenever $H_{z^{\mathsf{T}}z} = 0$ in which case

$$H_c(z) \;=\; \begin{bmatrix} H_{z^*z} & 0 \\ 0 & (H_{z^*z})^{\mathsf{T}} \end{bmatrix} \qquad (2M \times 2M) \tag{B.38}$$

For example, as shown in the calculation leading to (B.36), block diagonal Hessian matrices, $H_c(z)$ or $H_c(u)$, arise when $g(z)$ is quadratic in z. Such quadratic functions are common in design problems involving mean-square-error criteria in adaptation and learning — see, e.g., expression (2.63) in the body of the text.

■

End of Lecture

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.