UNIVERSIDAD DE LA FRONTERA

FACULTAD DE INGENIERÍA Y CIENCIAS



NUEVOS MÉTODOS DE LINEALIZACION PARA LA ECUACIÓN DE RICHARDS

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor en Ciencias mención Matemática

AUTOR: GUILLERMO ALEXIS ALBUJA PROAÑO

TEMUCO-CHILE

2021

Guillermo Alexis Albuja Proaño Tutor: Dr. Andres Ávila Barrera

2000 Mathematics Subject Classification 65M12, 35K15, 35K65, 65M08, 76S05, 65M22

Keywords and phrases: flujo en medios porosos, ley de Darcy, ecuación de Richards, esquemas de linealización, método de Newton, método de Picard, método L-scheme, método de elementos finitos de Galerkin.

INFORME DE APROBACIÓN

TESIS DE DOCTORADO FACULTAD DE INGENIERIA Y CIENCIAS UNIVERSIDAD DE LA FRONTERA

Se informa a la Vicerrectoría de Postgrado de la Universidad de La Frontera que la Tesis de Doctorado presentada por el candidato

GUILLERMO ALEXIS ALBUJA PROAÑO

ha sido aprobada por la Comisión de Evaluación de la tesis como requisito parcial para optar al grado de Doctor en Ciencias mención Matemáticas en el examen de Defensa de Tesis rendido el día xx/xx/2021

Nombre 1 (XXX)

Tutor: Dr. Andres Ávila Barrera

Nombre 2 (XXX)

Nombre 3 (XXX)

Temuco, Chile, XXXX 2021

DEDICATORIA

Dedico este trabajo en primer lugar y ante todo a Dios, por haberme dado la vida y permitirme el haber llegado hasta este momento tan importante de mi formación.

A mi querida familia que son mi esposa Carlita y mis hijas Alexita y Doménica, por ser el pilar más importante en mi vida, por demostrarme siempre su cariño y apoyo incondicional en todos los momentos de mi vida.

A mi padres Ángel y Judith, con mucho amor y cariño, les dedico todo mi esfuerzo, en reconocimiento a todo el sacrificio puesto para que yo pueda estudiar, se merecen esto y mucho más, a mis hermanos quienes fueron un gran apoyo emocional durante el tiempo en que escribía esta tesis.

Guillermo Albuja

AGRADECIMIENTOS

Agradezco principalmente a Dios por haberme dado la vida y permitirme culminar mis estudios junto a todos mis seres queridos.

Me gustaría agradecer a la UNIVERSIDAD CENTRAL DEL ECUADOR por darme la oportunidad de estudiar y por financiar mis estudios de doctorado, también quisiera expresar mi agradecimiento a la UNIVERSIDAD DE LA FRONTERA y de manera especial a todo el departamento de Matemática y Estadística a mis profesores ya que todos han aportado con un granito de arena a mi formación.

Quiero además expresar mi sincero agradecimiento al Dr.Andrés Ávila Barrera por su esfuerzo y dedicación, gracias a sus conocimientos, experiencia, paciencia y motivación he logrado terminar mis estudios con éxito, muchas gracias Andrés por tus consejos, enseñanzas y más que todo por tu amistad.

Son muchas las personas que han formado parte de mi vida profesional a las que me encantaría agradecerles su amistad, consejos, apoyo, ánimo y compañía en los momentos más difíciles de mi vida. Algunas están aquí conmigo y otras en mis recuerdos y en mi corazón, sin importar en donde estén quiero darles las gracias por formar parte de mí, por todo lo que me han brindado y por todas sus bendiciones.

Para ellos: Muchas gracias y que Dios los bendiga

RESUMEN Y ESTRUCTURA DE LA TESIS

Existen muchas aplicaciones importantes para la sociedad que involucran el flujo en medios porosos, por ejemplo la contaminación del agua y del suelo, el almacenamiento de CO_2 , la gestión de desechos nucleares, la recuperación mejorada de petróleo, por nombrar algunas. Los modelos matemáticos para el flujo multifásico en medios porosos involucran ecuaciones diferenciales parciales no lineales acopladas en dominios geométricamente complejos y que pueden variar en el tiempo, pues la mayoría de las aplicaciones que mencionamos son evolutivas. En este trabajo consideramos un caso particular de flujo bifásico que es usado para modelar el flujo de agua en el suelo.

El flujo agua a través de medios porosos está matemáticamente descrito por la Ecuación de Richards. Esta es una ecuación diferencial doblemente no lineal degenerada que no posee solución analítica, sin embargo existen algunos trabajos que dan soluciones analíticas para ciertos casos particulares ver por ejemplo [1]. La ecuación Richards es difícil de resolver numéricamente, más aún la no linealidad y la degeneración hacen que el diseño de esquemas numéricos para este problema sea una tarea desafiante.

El modelado matemático y la simulación numérica son herramientas poderosas y bien reconocidas para predecir el flujo en medios porosos, y por lo tanto permiten comprender y finalmente resolver los problemas mencionados anteriormente. Existen una variedad de métodos numéricos para resolver la Ecuación de Richards pero según la literatura los métodos implícitos son los que dan mejores resultados ya que los esquemas obtenidos permiten simular el problema degenerado. Sin embargo estos producen problemas no lineales que deben ser resueltos mediante métodos de linealización. En cuanto a la discretización espacial, hay varias opciones posibles, como son: diferencias finitas, elementos finitos de Galerkin, elementos finitos mixtos, volúmenes finitos entre los principales.

Cuando se usan esquemas implícitos para resolver la Ecuación de Richards, su rendimiento depende en gran medida de la robustez y eficiencia del método de linealización que estemos utilizando. Es así que a pesar de la intensa actividad investigadora de las últimas décadas, todavía existe una gran necesidad de desarrollar métodos de linealización robustos y eficientes para la Ecuación de Richards. Los principales métodos de linealización que se han usado para está ecuación son: método de *Newton-Raphson*, método de *Picard*, método de *Picard Modificado* y más recientemente el *L-scheme*. El método de *Newton* es cuadráticamente convergente y ha sido aplicado con mucho éxito a la Ecuación de Richards, sin embargo tiene algunos inconvenientes, es sólo localmente convergente y necesita el cálculo de algunas derivadas. El método *Picard*, aunque ampliamente utilizado, no es una buena opción cuando se aplica a la Ecuación de Richards pues no tiene un buen rendimiento y es menos robusto que el esquema de *Newton* [2]. El método *Picard Modificado* aparece como una mejora al método de *Picard*, este es solo linealmente convergente, pero es más robusto que el método de *Newton* [3,4]. El *L-scheme* es un método muy robusto y no necesita el cálculo de derivadas, más aún su convergencia global está demostrada en varios trabajos, sin embargo es solo linealmente convergente .

Esta tesis propone nuevos esquemas de linealización para la Ecuación de Richards. Los esquemas desarrollados son más robustos y eficientes que los ya existentes por lo que se convierten en una extraordinaria alternativa. Se demuestra rigurosamente la convergencia de los nuevos esquemas y se discute la convergencia de los otros métodos. Los métodos de linealización se prueban con varios ejemplos numéricos ilustrativos y se compara el rendimiento de los métodos propuestos con los esquemas existentes. Se estudia el rendimiento de los esquemas con respecto a su robustez, tiempo de CPU y orden de convergencia. Utilizamos un Esquema de Euler semi-implícito para discretizar el tiempo y el método elementos finitos para discretizar el espacio, sin embargo, estos métodos deberían funcionar para cualesquier discretización espacial escogida.

La tesis está dividida en cinco capítulos. En el primer capítulo se estudia el modelo matemático y sus implicaciones físicas, así como también hacemos una revisión de todos los trabajos que se han publicado sobre el tema hasta la fecha. En el segundo capítulo hacemos un estudio detallado de los principales métodos de linealización para la Ecuación de Richards, tomamos las ideas de varios trabajos para probar rigurosamente la convergencia de los esquemas y finalmente discutimos estos hallazgos teóricos. En el capítulo tres, buscamos una generalización del L-scheme probamos rigurosamente su convergencia y la usamos para construir una familia de esquemas de linealización de primer orden, robustos, eficientes y globalmente convergentes. Proponemos cinco esquemas que mejoran la convergencia del L-scheme uno de estos con convergencia de segundo orden, finalmente verificamos los resultados teóricos mediante tres ejemplos numéricos. En el cuarto capítulo proponemos tres nuevos esquemas de linealización para la Ecuación de Richards, se prueba teóricamente su convergencia, todos son robustos y tienen orden de convergencia superlineal pudiendo incluso en algunos casos llegar a ser cuadrática. El primer método es una combinación eficiente de los métodos de Newton y L-scheme, el segundo esquema está inspirado en el método de la secante y no usa las derivadas de las funciones involucradas y el último método combina el segundo método con el L-scheme. Los métodos se prueban en cuatro ejemplos numéricos ilustrativos en dos y tres dimensiones espaciales. El capítulo cinco está dedicado a las conclusiones, observaciones y recomendaciones finales.

Contenido

D	EDIO	CATORIA	vi			
AGRADECIMIENTOS v						
R	ESU	IMEN Y ESTRUCTURA DE LA TESIS	S viii			
IN	NDIC	CE DE FIGURAS	xiii			
IN	NDIC	CE DE TABLAS	xiii			
1	MO	ODELO MATEMÁTICO	1			
	1.1	Derivación de la Ecuación de Richards				
		1.1.1 Ecuación de Continuidad				
		1.1.2 Ley de Darcy-Buckingham				
		1.1.3 Ecuación de Richards				
		1.1.4 Modelo de Van-Genuchten				
		1.1.5 Hipótesis Matemáticas	6			
	1.2 Ecuación de Richards					
		1.2.1 Formas de la Ecuación de Richards				
		1.2.2 Existencia y unicidad de solución de	la Ecuación de Richards . 9			
	1.3	Marco Numérico				
		1.3.1 Discretización en el tiempo				
		1.3.2 Discretización en el Espacio				
2	MÉ	ÉTODOS DE LINEALIZACIÓN	13			
	2.1	Método de Newton				
		2.1.1 Esquema de Newton para la cuarta f	forma 15			
		2.1.2 Esquema de Newton para la forma n	nixta $\ldots \ldots \ldots \ldots \ldots \ldots \ldots 20$			
	2.2	2.2 Método de Picard Modificado				
		2.2.1 Esquema de Picard para la cuarta fo	orma			
2.3 Método <i>L</i> -scheme		Método L -scheme				
		2.3.1 L-esquema para la cuarta forma				
		2.3.2 L-scheme para la forma mixta				

		2.3.3 L-scheme cuando θ es Hölder continua	35			
	2.4	Esquemas de linealización mixtos	39			
3	UN	A FAMILIA DE NUEVOS ESQUEMAS DE LINEALIZACIÓN	40			
	3.1	Discretización del tiempo	41			
	3.2	Linealización y estimación del Error	42			
	3.3	Esquema de linealización	45			
		3.3.1 L-scheme y L2-scheme	45			
		3.3.2 Esquemas $DGLS$ y GLS	45			
		3.3.3 Esquema MNS	47			
		3.3.4 Esquemas <i>MDGLS</i> y <i>MGLS</i>	48			
	3.4	Experimentos Numéricos	50			
		3.4.1 Ejemplo 1	50			
		3.4.2 Ejemplo 2	52			
		3.4.3 Ejemplo 3	57			
	3.5	Conclusiones	62			
4	NUEVOS ESQUEMAS DE ORDEN SUPERLINEAL					
	4.1	Esquema L-Newton	65			
	4.2	Esquema Type-Secant	68			
	4.3	Esquema L -Secant	70			
	4.4	Experimentos Numéricos	72			
		4.4.1 Ejemplo 1	72			
		4.4.2 Ejemplo 2	73			
		4.4.3 Ejemplo 3	81			
		4.4.4 Ejemplo 4	84			
	4.5	Conclusión	87			
5	CO	NCLUSIONES Y RECOMENDACIONES	88			
	5.1	Conclusiones	89			
	5.2	Recomendaciones	90			

xii

Lista de Figuras

1.1	Ejemplo del modelo de Van-Genuchten	5
3.1	Funciones L_1 y L_2 para los esquemas $DGLS$ y GLS	46
3.2	Número de iteraciones	55
3.3	Tiempo de CPU.	55
3.4	Evolución de \hat{L}_1^n para el esquema MNS con el paso de tiempo $h_t = 0.25$	
	y un tamaño de malla de $h = 0.027$	56
3.5	Tiempo de CPU para $h_t = 1$ con la condición inicial $\Psi_0 = -61.5.$	60
3.6	Tiempo de CPU para $h_t = 1$ con la condición inicial $\Psi_0 = -40$	61
4.1	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$.	78
$4.1 \\ 4.2$	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$.	78 78
4.1 4.2 4.3	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$ Evolucion de L_1^n para pasos de tiempo $h_t = 1$ con un tamaño de malla	78 78
4.1 4.2 4.3	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Evolucion de L_1^n para pasos de tiempo $h_t = 1$ con un tamaño de malla h = 0.02 Ejemplo 2.	78 78 79
 4.1 4.2 4.3 4.4 	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$ Evolucion de L_1^n para pasos de tiempo $h_t = 1$ con un tamaño de malla h = 0.02 Ejemplo 2	78 78 79 80
 4.1 4.2 4.3 4.4 4.5 	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Evolucion de L_1^n para pasos de tiempo $h_t = 1$ con un tamaño de malla h = 0.02 Ejemplo 2	78 78 79 80 83
 4.1 4.2 4.3 4.4 4.5 4.6 	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$ Evolucion de L_1^n para pasos de tiempo $h_t = 1$ con un tamaño de malla h = 0.02 Ejemplo 2 Evolución de λ_1^n con $h_t = 1$ y $h = 0.02$, para el Ejemplo 2 Evolución de λ_{k+1}^n con $h_t = 1$, $h = 0.3195$ y $\Psi_0 = -40$, (Ejemplo 3). Geometría y condiciones de borde del Ejemplo 4	78 78 79 80 83 84
$\begin{array}{c} 4.1 \\ 4.2 \\ 4.3 \\ 4.4 \\ 4.5 \\ 4.6 \\ 4.7 \end{array}$	Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$. Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$ Evolucion de L_1^n para pasos de tiempo $h_t = 1$ con un tamaño de malla h = 0.02 Ejemplo 2 Evolución de λ_1^n con $h_t = 1$ y $h = 0.02$, para el Ejemplo 2 Evolución de λ_{k+1}^n con $h_t = 1$, $h = 0.3195$ y $\Psi_0 = -40$, (Ejemplo 3). Geometría y condiciones de borde del Ejemplo 4 Perfiles de presión para $T = 20$, $h_t = 1$ y un tamaño de malla $h = 0.155$	78 78 79 80 83 84

Lista de Tablas

3.1	$\ \Psi_n - \Psi\ $ para el Ejemplo 1	51
3.2	$\ \nabla \Psi_n - \nabla \Psi\ $ para el Ejemplo 1	52
3.3	Número de iteraciones, tiempo de CPU y orden de convergencia para	
	$h_t = 0.25, 1, 5$ y cinco tamaños de mallas h diferentes	54
3.4	Número de iteraciones, tiempo de CPU y orden de convergencia para	
	los tamaños de malla 0.9428, 0.6248 y 0.3195, con pasos de tiempo	
	$h_t = 1, 10$, usando la condición inicial $\Psi_0 = -61.5$	58
3.5	Número de iteraciones, tiempo de CPU y orden de convergencia para	
	los tamaños de malla de $0.9428, 0.6248$ y 0.3195 , con pasos de tiempo	
	$h_t = 1, 10$, usando la condición inicial $\Psi_0 = -40.0$	59
4 1		
4.1	$\ \Psi_n - \Psi\ $ para el Ejemplo 1	74
4.2	$\ \nabla \Psi_n - \nabla \Psi\ $ para el Ejemplo 1	75
4.3	Número de iteraciones y tiempo de CPU para el Ejemplo 2	77
4.4	Iteraciones y tiempo de CPU para tamaños de malla de 0.9428, 0.6248,	
	y 0.3195 usando un paso de tiempo de $h_t = 1, 10$ para la condición	
	inicial $\Psi_0 = -61.5$ (Ejemplo 3)	82
4.5	Iteraciones y tiempo de CPU para tamaños de malla de 0.9428, 0.6248,	
	y 0.3195 usando un paso de tiempo de $h_t = 1,10$ para la condición	
	inicial $\Psi_0 = -40$ (Ejemplo 3).	83
4.6	Número de iteraciones y tiempo de CPU para tamaños de malla h de	
	$0.354, 0.155 \text{ y } 0.097 \text{ y pasos de tiempo } h_t \text{ de } 0.25, 1 \text{ y } 5.$ (Ejemplo 4).	85

Capítulo 1 MODELO MATEMÁTICO

1.1 Derivación de la Ecuación de Richards

En esta sección presentamos un modelo matemático de flujo derivado de las leyes de conservación de masa y de Darcy. El modelo resultante es la ecuación diferencial de Richards y tiene muchísimas aplicaciones en varias áreas de la ingeniería, por lo que es de vital importancia para la ciencia y la sociedad estudiar y entender este fenómeno. Algunas de sus aplicaciones son los flujos de aguas subterráneas, la dinámica de humedad en materiales de construcción, simulación de yacimientos de petróleo, en la agricultura permite modelar dinámica del flujo de agua en el suelo y su efecto significativo en el crecimiento de las plantas [5].

1.1.1 Ecuación de Continuidad

Un medio poroso es una matriz sólida indeformable con huecos que pueden o no estar interconectados llamados *poros*. Los poros permiten el paso de un fluido a través del medio poroso. La porosidad del medio representa la fracción del espacio total que puede ser ocupado por el fluido. En este contexto podemos definir la porosidad η $[L^3/L^3]$ como la relación entre el volumen total V_T y el volumen de los poros V_P

$$\eta = \frac{V_P}{V_T}.$$

Resulta evidente que $0 \le \eta \le 1$. Dado que el fluido circula a través de los poros, el volumen de fluido en el interior del medio poroso V_F toma un valor inferior al volumen de los poros. Definimos el contenido de humedad $\theta [L^3/L^3]$ como la fracción total de volumen del medio poroso que está ocupada por el fluido

$$\theta = \frac{V_F}{V_T}.$$

Se deduce inmediatamente que $0 \le \theta \le \eta \le 1$. La saturación del fluido se define como

$$S_F = \frac{V_F}{V_P}.$$

de donde se tiene que

$$\theta = \eta S_F.$$

Según Marinoschi (2010) [6] debemos considerar las siguientes hipótesis sobre el medio poroso y el fluido:

- (i) El medio poroso es isótropo, indeformable y homogéneo con porosidad constante.
- (ii) El fluido es incompresible y no tiene reacciones químicas con el medio poroso.
- (iii) El flujo no tiene histéresis.

Consideramos que no existe una distribución uniforme de fluido en el interior del medio poroso y esta varia en el tiempo, por lo que el contenido de humedad es distinto en cada punto del espacio y en cada instante de tiempo. Por esta razón se considera el contenido de humedad como función del espacio y del tiempo. También se conoce que el contenido de humedad es función de la cabeza de presión Ψ [L], mediante la conocida curva de retención de humedad. Esta relación será tratada con mayor detalle posteriormente, pero se encuentra bien documentada en [7,8].

Físicamente $\theta \geq \theta_r$, donde θ_r es una constante positiva a la que llamaremos contenido de humedad residual y $\theta \leq \theta_s < \eta$, donde θ_s es una constante denominada contenido de humedad saturada. Cuando $\theta = \theta_s$ se dice que es un **flujo saturado**, si por el contrario $\theta < \theta_s$, entonces diremos que se trata de un **flujo no saturado**, es decir, los poros no se encuentran completamente llenos de fluido sino que además contienen aire.

Sea $\Omega \subset \mathbb{R}^d$ con $d \in \{1, 2, 3\}$ un dominio abierto acotado conexo que representa el medio poroso incluyendo los poros, y (0, T) un intervalo finito de tiempo, donde Tes una constante positiva fija. Definimos el contenido de humedad como la función $\theta : \Omega \times (0, T) \rightarrow [\theta_r, \theta_s]$, donde $\theta(\vec{x}, t)$ es el contenido de humedad en un punto del espacio $\vec{x} \in \Omega$ y un tiempo t determinado.

Sea $\omega \subset \Omega$ un volumen de control fijo pero arbitrario, entonces el volumen total de fluido contenido en ω en un tiempo t está dado por:

$$V(t) = \int_{\omega} \theta(\vec{x}, t) dx.$$

Puesto que el fluido es incompresible, el principio de conservación obliga a que el volumen total de fluido en el interior de ω se mantenga constante en el tiempo,

$$\frac{dV(t)}{dt} = \frac{d}{dt} \int_{\omega} \theta(\vec{x}, t) dx = 0.$$

Usando en la ecuación anterior, el teorema de transporte de Reynolds y el teorema de la divergencia de Gauss se tiene que:

$$\int_{\omega} \left[\frac{\partial \theta}{\partial t} + \operatorname{div}(\vec{q}) \right] dx = 0,$$

donde $\vec{q} [LT^{-1}]$ es el flujo. Como ω es arbitrario se deduce que

$$\frac{\partial \theta}{\partial t} + \operatorname{div}(\vec{q}) = 0 \text{ sobre } \Omega \times (0, T).$$

De manera más general si consideramos la existencia de fuentes o sumideros se tiene la ecuación siguiente:

$$\frac{\partial \theta}{\partial t} + \operatorname{div}(\vec{q}) = S \text{ sobre } \Omega \times (0, T),$$
(1.1)

donde $S : \Omega \times (0,T) \to \mathbb{R}$ es una función que representa las posibles fuentes o sumideros. Para mayores detalles sobre la ecuación de continuidad nos podemos referir a Jacques y Delleur (2006) [9].

1.1.2 Ley de Darcy-Buckingham

El punto más importante para el modelo matemático de flujo en medios porosos es la ley de Darcy formulada en 1856 por el ingeniero francés Henry Darcy [9]. La ley fue originalmente introducida para flujo en medios porosos saturados, y posteriormente extendida para medios porosos no saturados por Buckingham en 1907 [10], y describe las características del movimiento de un fluido a través de un medio poroso continuo, homogéneo e isótropo y está descrita por:

$$\vec{q} = -K(\theta)\nabla H,$$

donde \vec{q} , $[LT^{-1}]$ es el flujo definido como el caudal por unidad de área del medio poroso, ∇H es el gradiente hidráulico adimensional y representa la pérdida o cambio de potencial hidráulico por unidad de longitud medida en el sentido del flujo, H = $\Psi + z$ [L] es el potencial hidráulico, donde Ψ [L] es la cabeza de presión y z [L] es la coordenada vertical positiva, $K(\theta)$ [LT^{-1}] es la conductividad hidráulica del medio poroso y representa la mayor o menor facilidad con la que el medio poroso deja pasar el fluido. θ también es función de Ψ y esta relación se encuentra dada a través de la denominada curva de retención de humedad. Por lo tanto la conductividad hidráulica K también es función de Ψ , es así que podemos escribir alternativamente $K(\theta)$ o $K(\Psi)$ o $K(\theta(\Psi))$. La Ley de Darcy-Buckingham está dada por:

$$\vec{q} = -K(\Psi)\nabla(\Psi + z)$$
 sobre $\Omega \times (0, T)$. (1.2)

Es importante notar que la conductividad hidráulica K depende de la cabeza de presión Ψ pero a través del contenido de humedad θ . Además debemos tener en cuenta que el fluido considerado en la Ley de Darcy-Buckingham y a través de este documento es incompresible y no viscoso.

1.1.3 Ecuación de Richards

El resultado de reemplazar la Ley de Darcy-Buckinghan (1.2) en la ecuación de continuidad (1.1) produce la famosa Ecuación de Richards. Esta fue propuesta por Richards en 1931 y describe el movimiento de un fluido a través de un medio poroso [11]

$$\frac{\partial \theta(\Psi)}{\partial t} - \operatorname{div}(K(\Psi)\nabla(\Psi + z)) = S \text{ sobre } \Omega \times (0, T).$$
(1.3)

En la zona saturada, las funciones θ y K son constantes y $\Psi \ge 0$. Por el contrario si el flujo es no saturado, θ y K son funciones no lineales de Ψ y además $\Psi < 0$, esto significa que la Ecuación de Richards es una ecuación diferencial no lineal. Además la ecuación degenera cuando $K(\theta(\Psi)) \rightarrow 0$, que es el caso de *difusión lenta*, o cuando θ es constante que se llama *difusión rápida*. La región de degeneración depende de la saturación del medio, por lo que estas regiones no son conocidas apriori y pueden variar en el tiempo y en el espacio. En este trabajo nos concentraremos especialmente en el caso de difusión rápida, donde es típica la degeneración elíptica-parabólica y la baja regularidad de las soluciones [12]. La no linealidad y la degeneración hacen que el diseño y análisis de los esquemas numéricos para la Ecuación de Richards sean tareas desafiantes.

Varios estudios experimentales han demostrado la dependencia no lineal que tienen las funciones θ y K de la cabeza de presión Ψ , por lo que varios modelos empíricos y semiempíricos se han propuesto, la mayoría de estos consideran una sola capa de suelo y se modelan a través de funciones no lineales continuas y no decrecientes. Sin embargo, existen trabajos donde se consideran modelos con menos regularidad, como por ejemplo discontinuidades que se producen cuando se consideran dos capas de suelo [13]. Uno de los modelos más populares fue propuesto por Van-Genuchten 1980 [7].

1.1.4 Modelo de Van-Genuchten

El modelo de Van-Genuchten produce un buen ajuste a las curvas de datos experimentales por lo que es usado frecuentemente para comparar nuevos modelos que han sido propuestos por investigadores, así como también es comúnmente utilizado para mostrar la bondad de los nuevos esquemas numéricos desarrollados, ver por ejemplo [2–4, 14–24]. Este modelo representa la curva de retención de humedad de la manera siguiente:

$$\theta(\Psi) = \begin{cases} \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha |\Psi|)^n]^m}, & \text{si } \Psi \le 0, \\ \theta_s, & \text{si } \Psi > 0, \end{cases}$$
(1.4)

donde θ_s es el contenido de humedad saturado, definido como el máximo contenido volumétrico de fluido que un suelo puede tener, se conoce además que θ_s no es igual a la porosidad del suelo, de hecho es un 5% a 10% más pequeño ya que existe aire atrapado o disuelto en el fluido. El contenido de humedad residual θ_r es la cantidad



Figura 1.1: Ejemplo del modelo de Van-Genuchten

de fluido en el suelo que no interviene en el flujo de la fase liquidada y toma valores estrictamente positivos, puesto que los poros pueden estar bloqueados o la absorción dada por la fase sólida es demasiado fuerte. Así, θ_s y θ_r son dos constantes positivas tales que $0 < \theta_r \leq \theta(\Psi) \leq \theta_s < +\infty$ para todo $\Psi \in \mathbb{R}$ y por lo general son obtenidas usando técnicas estadísticas en los datos experimentales. Los parámetros α , n, y mson constantes propias del suelo que afectan directamente a la forma de la curva, más aún α es un parámetro empírico cuya inversa es normalmente asociada al valor de la presión de entrada de aire, el parámetro m se toma como $m = 1 - \frac{1}{n}$ o $m = 1 - \frac{2}{n}$.

La conductividad hidráulica en función de la cabeza de presión que la notaremos como $K(\Psi)$ está dada por:

$$K(\Psi) = \begin{cases} K_s \frac{\left[1 - (\alpha |\Psi|)^{n-1} (1 + (\alpha |\Psi|)^n)^{-m}\right]^2}{\left[1 + (\alpha |\Psi|)\right]^{\frac{m}{2}}}, & \text{si } \Psi \le 0, \\ K_s, & \text{si } \Psi > 0, \end{cases}$$
(1.5)

donde la constante K_s representa la conductividad hidráulica cuando el medio poroso está saturado. Unos cálculos simples nos llevan a concluir que en el modelo de Van-Genuchten, la curva de retención de humedad y la conductividad hidráulica son funciones continuas, monótonas no decrecientes, positivas y acotadas. La función $\theta(\Psi)$ es derivable para n > 1 y si $n \leq 1$ la función es derivable en todos los puntos excepto en 0. La conductividad hidráulica es derivable para n > 2 y si $n \leq 2$ la derivabilidad falla en 0. Además la elección de n > 2 implica que las funciones $\theta(\Psi)$ y $K(\Psi)$ son Lipschitz continuas. Los valores de los parámetros de Van-Genuchten para varios tipos de suelo pueden consultarse en [25]. Observe que en los dos modelos descritos anteriormente cuando $\Psi > 0$, es decir, cuando el flujo es saturado, las dos funciones son constantes. En la Figura 1.1 se puede distinguir las gráficas de estos modelos.

Otro modelo bastante popular es el propuesto por Haverkamp en 1977 [8], este ha

sido utilizado en varios trabajos [3, 15, 26–30]. Este modelo está dado por:

$$\theta(\Psi) = \begin{cases} \theta_r + \frac{\alpha(\theta_s - \theta_r)}{\alpha + |\Psi|^{\beta}}, & \text{si} \quad \Psi \le 0, \\ \theta_s, & \text{si} \quad \Psi > 0, \end{cases}$$
(1.6)

$$K(\Psi) = \begin{cases} K_s \frac{A}{A+|\Psi|^{\gamma}}, & \text{si} \quad \Psi \le 0, \\ K_s, & \text{si} \quad \Psi > 0, \end{cases}$$
(1.7)

donde β , A y γ son parámetros adimensionales de ajuste de las curvas.

1.1.5 Hipótesis Matemáticas

En esta parte vamos a especificar de manera general las propiedades matemáticas de las funciones hidráulicas involucradas. Una referencia muy importante donde se encuentran muy bien documentadas estas propiedades es el libro de Marinoschi (2010) [6] en las páginas 20,21 y 177.

Hipótesis 1.1.1. Asumimos las siguientes hipótesis sobre θ y K :

(i) La curva de retención de humedad es una función estrictamente positiva, acotada y continua

$$\begin{array}{rcl} \theta : & \mathbb{R} & \longrightarrow [\theta_r, \theta_s] \\ & \Psi & \longrightarrow \theta(\Psi) \end{array}$$

donde θ_r y θ_s son dos constantes tales que $0 < \theta_r < \theta_s < 1$. Además θ cumple con las siguientes propiedades:

- $\theta(\Psi) = \theta_s$ para todo $\Psi \ge 0$ y $\theta(\Psi) < \theta_s$ para todo $\Psi < 0$.
- El contenido de humedad restringido al intervalo (−∞,0) es una función monótona creciente y dos veces diferenciable.
- $\lim_{\Psi \to 0} \theta(\Psi) = \theta_s.$
- (ii) La función de capacidad específica se define como la derivada del contenido de humedad respecto a la cabeza de presión y es una función acotada, positiva y no necesariamente continua en cero

$$\frac{d\theta}{d\Psi} = C : \mathbb{R} \longrightarrow [0, C_0]$$
$$\Psi \longrightarrow C(\Psi)$$

donde C_0 es una constante positiva. C cumple con las siguientes propiedades

- $C(\Psi) = 0$ para todo $\Psi > 0$.
- La capacidad específica restringida al intervalo (−∞, 0) es una función no negativa, acotada y diferenciable.
- Existe un único $\Psi_0 \in (-\infty, 0)$ tal que:

$$C(\Psi_0) = \max_{\Psi \in \mathbb{R}} C(\Psi)$$

(iii) La conductividad hidráulica se define como una función estrictamente positiva acotada y continua

$$\begin{array}{rcl} K : & \mathbb{R} & \longrightarrow [K_r, K_s] \\ & \Psi & \longrightarrow K(\Psi) \end{array}$$

donde K_r y K_s son números reales fijos, tales que $0 < K_r < K_s < +\infty$. Esta función cumple con las siguientes propiedades:

- $K(\Psi) = K_s$ para todo $\Psi > 0$ y $K(\Psi) < K_s$ para todo $\Psi < 0$.
- La conductividad hidráulica restringida al intervalo $(-\infty \ 0)$ es una función monótona creciente y dos veces diferenciable.
- $\lim_{\Psi \to 0} K(\Psi) = K_s.$

.

Es fácil verificar que los modelos de Van-Genuchten y Haverkamp cumplen con las hipótesis 1.1.1

1.2 Ecuación de Richards

Sea $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$ un abierto conexo acotado con frontera $\partial \Omega$ Lipschitz continua, definimos $\Omega_T = \Omega \times (0, T)$ para una constante fija T > 0. Consideramos la siguiente ecuación diferencial parcial cuasi lineal degenerada elíptica-parabólica:

$$\begin{cases} \frac{\partial \theta(\Psi)}{\partial t} - \operatorname{div} \left[K(\Psi) \nabla \left(\Psi + z \right) \right] = S & \text{sobre } \Omega_T \\ + \operatorname{Condiciones \ iniciales} \\ + \operatorname{Condiciones \ de \ borde} \end{cases}$$
(1.8)

1.2.1 Formas de la Ecuación de Richards

En la literatura se puede encontrar la Ecuación de Richards principalmente en cuatro formas, las tres primeras se las puede consultar en los trabajos de Celia et al. (1990) [3], Caviedes et al. (2013) [19], Berardia et al. (2016) [29], Marinoschi (2010) [6], y la cuarta forma se la puede encontrar en los trabajos de Radu et al. (2004) [31], Pop et al. (2004) [32], Radu et al (2008) [33]. A continuación se presentan de forma resumida.

Ecuación de Richards en su forma mixta:

Se expresa en términos de la cabeza de presión y el contenido de humedad, y es la ecuación (1.8).

Ecuación de Richards en términos en presión (basada- Ψ):

Se expresa en términos de la cabeza de presión Ψ . Para obtenerla se usa la regla de la cadena en el primer término de la ecuación (1.8)

$$\frac{\partial \theta(\Psi)}{\partial t} = \frac{\partial \theta(\Psi)}{\partial \Psi} \frac{\partial \Psi}{\partial t} = C(\Psi) \frac{\partial \Psi}{\partial t},$$

donde la función $C(\Psi)$ se conoce como función de capacidad específica, luego la ecuación deseada es:

$$C(\Psi)\frac{\partial\Psi}{\partial t} - \operatorname{div}(K(\theta(\Psi))\nabla(\Psi+z)) = S.$$
(1.9)

Ecuación de Richards en términos del contenido de humedad (basada- θ):

Se escribe en términos del contenido de humedad θ , por lo que la ecuación es válida solamente para la zona no saturada. En esta zona suponemos que la función $\theta(\Psi)$ es una biyección, por lo tanto tiene sentido escribir $\Psi(\theta)$. Usando la regla de la cadena se obtiene

$$\nabla \Psi(\theta) = \frac{\partial \Psi}{\partial \theta} \nabla \theta$$

por lo tanto se tiene:

$$K(\theta)\nabla\Psi(\theta) = K(\theta)\frac{\partial\Psi}{\partial\theta}\nabla\theta = D(\theta)\nabla\theta,$$

donde a $D(\theta)$ se conoce como difusividad. Observe que la conductividad hidráulica $K(\Psi(\theta))$ la notamos simplemente como $K(\theta)$, por lo que la ecuación (1.8) se puede escribir como

$$\frac{\partial \theta}{\partial t} - \operatorname{div}(D(\theta)\nabla(\theta)) - \frac{\partial K(\theta)}{\partial z} = S.$$
(1.10)

La ecuación (1.10) se encuentra descrita en términos de la variable conservada θ y no se puede resolver cuando el flujo se satura puesto que θ y K se vuelven constantes y los términos del lado izquierdo de la ecuación se anulan. Más aún, es bien conocido que esta ecuación en condiciones de heterogeneidad produce soluciones discontinuas.

Está claro que esta ecuación no puede representar adecuadamente el fenómeno físico, pues en un instante de tiempo determinado el medio porosos puede presentar regiones donde el flujo se satura y otras donde el flujo es no se satura. Esta es una de las razones por la que (1.10) no ha despertado mucho interés en la comunidad científica, de hecho solo hemos encontrado una referencia de los últimos 10 años [1], donde se obtiene una solución analítica para un caso particular.

Cuarta forma de la Ecuación de Richards:

La cuarta forma resulta de aplicar la transformada Kirchhoff, con esto se consigue combinar las dos principales no linealidades en una sola. De acuerdo con [12,31–33] la transformada esta dada por:

$$\begin{aligned} \mathcal{K} : & \mathbb{R} & \to & \mathbb{R} \\ & \Psi & \to & \mathcal{K}(\Psi) = \int_{0}^{\Psi} K(\theta(s)) ds. \end{aligned}$$

Si definimos la función $u = \mathcal{K}(\Psi)$ se tiene inmediatamente que $\nabla u = K(\theta(\Psi))\nabla\Psi$. Por otro lado de las propiedades de K y θ se deduce que la función \mathcal{K} es invertible, es así que tomamos $b = \theta \circ \mathcal{K}^{-1}$ de donde se tiene que $\theta(\Psi) = b(u)$, finalmente reemplazando en la forma mixta de la Ecuación de Richards se tiene:

$$\frac{\partial b(u)}{\partial t} - \operatorname{div}(\nabla u + K(b(u))\vec{e}_z) = S, \qquad (1.11)$$

donde \vec{e}_z es el vector unitario en la dirección z.

1.2.2 Existencia y unicidad de solución de la Ecuación de Richards

La existencia y unicidad de solución de la Ecuación de Richards para diversas condiciones de borde puede ser consultado en Marinoschi (2010) [6],Van Duyn et al. (1982) [34], Alt (1983) [12], Kačur (1990) [35], y Otto (1996) [36]. Sin, embargo el trabajo más importante es el de Alt (1983) [12], que estudia un sistema de ecuaciones diferenciales no lineales elíptico-parabólicas que constituye un problema muy general donde la de Richards es un caso particular.

1.3 Marco Numérico

A la largo de esta sección describiremos brevemente los diferentes esquemas numéricos que se han usado para resolver la Ecuación de Richards. Hemos dividido esta sección en dos partes, los esquemas para discretizar el tiempo y los esquemas para discretizar el espacio dejando los métodos de linealización para tratarlos con mayor detalle en el Capítulo 2.

1.3.1 Discretización en el tiempo

El esquema más popular para discretizar el tiempo es el Esquema de Euler, ver por ejemplo [3, 4, 13-15, 19, 21, 23, 26, 28, 31-33, 37-43]. Aún cuando las dos formas de la Ecuación de Richards, basada- Ψ (1.9) y la forma mixta (1.8), son equivalentes,

sus aproximaciones numéricas cuando se usa un Esquema de Euler pueden conducir a resultados completamente diferentes. Esto se muestra en los siguientes trabajos: Celia et al. (1990) [3] para un Esquema de Euler implícito y una discretización espacial de diferencias finitas y elementos finitos 1D, Lehmann et al (1998) [14] para un Euler Implícito y elementos finitos 1D, y Caviedes D. et al. [19] que usa los esquemas de Euler implícito y explicito para el tiempo y el método de volúmenes finitos 1D para el espacio. Estos autores demuestran que la discretización con un Esquema de Euler, de la formulación basada- Ψ , arroja resultados pobres caracterizados por un gran error en el balance de masa y estimaciones erróneas de la profundidad de infiltración, en contraste con la discretización de la forma mixta que produce un correcto balance de masa. Una discretización de Euler aplicada a (1.9) solo consigue soluciones aceptables para pasos de tiempo muy pequeños. Esto ocurre por la incapacidad del esquema para conservar la propiedad de balance de masa, aún cuando debemos señalar que un correcto balance de masa es una condición necesaria pero no suficiente para obtener una solución correcta según lo demuestra Lehmann F. (1998) [14]. Es evidente que (1.11) también cumple con esta propiedad de conservación cuando se usa los esquemas de Euler, pues puede ser vista como un caso particular de la forma mixta.

Existen varios trabajos que han sugerido adaptaciones temporales que preservan el equilibrio de masa para la forma basada- Ψ . Celia et al. (1990) [3] propone utilizar una aproximación de segundo orden de precisión en tres niveles para la derivada temporal de la cabeza de presión Ψ

$$\frac{\partial \Psi_{k+1}}{\partial t} \approx \frac{3\Psi_{k+1} - 4\Psi_k + \Psi_{k-1}}{h_t}$$

Esta aproximación produce mejoras significativas en el balance de masa y precisión de la solución. Tocci et al. (1997) [44] propone tomar pasos de tiempo variables en lugar de pasos de tiempo constantes, para lo cual hace uso del método de líneas obteniendo soluciones precisas y con buenas propiedades de equilibrio de masa. Kuraz et al. (2010) [16] también propone pasos de tiempo adaptativos, para lo cual sugiere el uso de un algoritmo que lo denomina *Retention Curve Zone Approach* el cual controla el error de aproximación de:

$$\frac{\Delta \theta(\Psi)}{\Delta t} \thickapprox C(\Psi) \frac{\Delta \Psi}{\Delta t}$$

El Esquema de Euler completamente explícito ha sido usado en [13,19]. Además de las desventajas ya conocidas de estos métodos es importante señalar que el esquema solo permite simular en la zona no saturada, lo cual es una gran desventaja [19]. Según List et al. (2016) [4], la mejor opción para la discretización temporal es un Esquema de Euler implícito en (1.8) o (1.11), esto se debe a dos razones: la necesidad de una discretización estable que permita pasos de tiempo largos y la baja regularidad de la solución que no soporta cualquier esquema de alto orden. La forma mixta y la cuarta forma de la ecuación de Richards con una discretización de Euler poseen las ventajas de la forma basada en Ψ y de la forma basada en θ , es decir, producen una discretización con una correcta propiedad de conservación de masa, además que proporcionan soluciones continuas a través de condiciones heterogéneas, incluso en transmisiones de régimen de no saturado a saturado.

Otros esquemas de tiempo han sido propuestos y estudiados por varios autores, en [2, 23] se usa el esquema de Crank-Nicolson. En [22] se proponen esquemas de tiempo implícitos mucho más generales que el Esquema de Euler implícito para pasos de tiempo constantes y variables. En [16, 21, 45–48] se usa el Esquema de Euler combinado con varias técnicas de paso de tiempo adaptativo.

1.3.2 Discretización en el Espacio

Las aproximaciones típicas que se aplican al dominio espacial son los esquemas de Diferencias Finitas, Elementos Finitos, Elementos Finitos Mixtos y Volúmenes Finitos. En [3] se estudia comparativamente los métodos de Diferencias Finitas y Elementos Finitos, en este trabajo se concluye que las aproximaciones por el Método de Elementos Finitos podrían sufrir oscilaciones, sin embargo este esquema ha sido muy utilizado en los últimos tiempos [23, 43, 49].

Diferencias Finitos

El Método de Diferencias Finitas (MDF) fue usado en la Ecuación de Richards en algunos trabajos entre estos podríamos citar [3, 29, 44, 45] y más recientemente en [29, 30, 50]. En [29] se presenta el MDF conjuntamente con una técnica numérica basada en el método de líneas (MoL).

Elementos Finitos

El Método de Elementos Finitos (MEF) es un método numérico general usado para la aproximación de soluciones de ecuaciones diferenciales parciales muy complejas. El MEF se basa fuertemente en la formulación variacional o débil del problema de valores de contorno. Para los detalles del método, su implementación y aplicaciones podemos referirnos a [51,52]. Existen varios trabajos que usan Elementos Finitos de Galerkin en el espacio para resolver numéricamente la Ecuación de Richards, En [3,4,14–16,46,53] se usa el MEF conjuntamente con un Esquema de Euler implícito en la forma mixta (1.8). En [3,14,16,23,47] se usa este método con un Esquema de Euler implícito para el tiempo en la ecuación basada- Ψ . En [42] se usó el MEF en la Ecuación de Richards luego de aplicar la transformada de Kirchhoff (1.11).

Elementos Finitos Mixtos

El Método de Elementos Finitos Mixtos (MEFM) se usa en la Ecuación de Richards por lo general empleando los elementos Raviart-Thomas y el espacio de funciones

constantes a trozos. En la mayoría de los trabajos consultados se ha usado en (1.11) [31–33, 41, 43, 54]. En [38, 39] se usa MEFM en la forma basada- Ψ (1.9).

Volúmenes Finitos

El Método de Volúmenes Finitos es útil para la simulación numérica de leyes de conservación como es el caso de la Ecuación de Richards. Existen varias referencias sobre esquemas numéricos de volúmenes finitos para la ecuación de Richards, ver por ejemplo [13, 17-19, 21, 22, 24, 26-28, 48]. La mayoría usa un Esquema de Euler implícito para el tiempo, sin embargo hay trabajos como [13, 19] que usan un Esquema de Euler explicito. En [13, 17, 19, 21, 22, 26, 48] se usa en la forma mixta mientras que en [19, 21, 24] se usa en la forma basada- Ψ y en [27] en la cuarta forma de la Ecuación de Richards.

Un artículo muy interesante es el de Zambra et al. (2012) [18], donde se construye un esquemas de volúmenes finitos de alto orden de precisión en el espacio y tiempo. En Eymard et al. [26] de prueba la convergencia del método de volúmenes finitos para la Ecuación de Richards usando el Teorema de Kolmogorov sobre subconjuntos de $L^2(\Omega)$ relativamente compactos. También Pop et al. [55] analizan la convergencia para una discretización de volúmenes finitos en una ecuación de flujo en medios porosos.

Estamos principalmente interesados en la forma mixta de la Ecuación de Richards aunque no descartamos completamente la cuarta forma. Algunos de los resultados teóricos estudiados en el capítulo siguiente solo se pueden obtener para esta formulación, pero en los capítulos tres y cuatro usaremos únicamente la forma mixta. En esta tesis no trataremos las ecuaciones basada- Ψ y basada- θ ya que presentan algunos problemas que fueron mencionados anteriormente. En cuanto a los métodos numéricos usaremos los esquemas de Euler para el tiempo y los Elementos Finitos de Galerkin para el espacio, sin embargo muchos de los esquemas aquí planteados son perfectamente aplicables a cualquier discretización espacial escogida.

Capítulo 2 MÉTODOS DE LINEALIZACIÓN

Para el lector que no esté familiarizado con los definiciones básicas sobre métodos de linealización y métodos iterativos recomendamos la lectura de [56–58]. En [56] se describe en forma breve muchos conceptos básicos, se estudian algunos de los métodos iterativos más populares y se describe cómo se puede aplicar estos métodos a la resolución de ecuaciones diferenciales no lineales elípticas y parabólicas. En [57] se definen muchos conceptos importantes usados frecuentemente durante el desarrollo de esta tesis tal como: método iterativo, orden de convergencia teórico y orden de convergencia experimental entre los más importantes.

De acuerdo con la definición dada en [56], un método iterativo se dice que es globalmente convergente si el método converge para cualquier solución inicial $x_0 \in D$ con la que se comienza el método iterativo (es decir, para valores iniciales arbitrarios en D), donde D es el dominio del problema. Si existe un $U \subset D$ abierto tal que el método es convergente para todo $x_0 \in U$, entonces se llama localmente convergente. En el último caso, U se denomina rango de la iteración. Finalmente de acuerdo con [4, 15] un método iterativo *localmente convergente* con rango de iteración U_1 se dice mas robusto que otro con rango U_2 si $U_2 \subset U_1$

Varios métodos de linelización han sido propuestos para la Ecuación de Richards por diversos autores [3, 4, 14, 15, 30, 43, 48, 49], así como también existen numerosos trabajos que discuten la convergencia de los métodos de linelización [2–4, 14, 15, 30, 32, 38, 41, 43, 48, 49].

Los métodos de linealización más populares y que han sido usados ampliamente en muchísimos trabajos son: el Esquema de Newton, Picard Modificado, L-scheme y ciertas combinaciones de estos métodos como son los esquemas de Picard/Newton y L-esquema/Newton. Es conocido que el método de Newton bajo ciertas hipótesis es cuadráticamente convergente, sin embargo tiene el inconveniente de ser solo locamente convergente, es decir, no está garantizada la convergencia del método para cualquier solución inicial con la que se comience el esquema iterativo y además necesita del cálculo de las derivadas de θ y K. Por otro lado el método de Picard Modificado y el L-scheme son solo linealmente convergentes pero en cambio son esquemas más

robustos, es decir, que convergen cuando el esquema de Newton no lo hace. En [14,41] se proponen combinaciones de estos esquemas con el método de Newton, es decir, se comienza haciendo unas pocas iteraciones con el método de Picard Modificado o el L-scheme para luego cambiar al método de Newton obteniendo esquemas robustos con orden de convergencia superlineal y en ocasiones cuadrática. Lastimosamente la convergencia de estos métodos tampoco está asegurada y además se sigue usando el método de Newton lo que supone seguir calculando derivadas.

En todo lo que se sigue utilizaremos las definiciones y propiedades de los siguientes espacios funcionales dados en [59] $L^2(\Omega)$, $H^1(\Omega)$, $H^0(\Omega)$, $H^{-1}(\Omega)$, $L^2(0,T;L^2(\Omega))$, $L^2(0,T;H^0(\Omega))$, $L^{\infty}(0,T;L^2(\Omega))$, y $L^2(0,T;H^{-1}(\Omega))$. Además denotamos por $\langle \cdot, \cdot \rangle$ el producto interior en $L^2(\Omega)$ o el par de dualidad entre $H^1_0(\Omega)$ y $H^{-1}(\Omega)$ según sea el caso, $\|\cdot\|$ la norma en $L^2(\Omega)$ y $\|\cdot\|_1$ la norma en el espacio $H^1(\Omega)$.

2.1 Método de Newton

Fue uno de los primeros métodos de linealización usado para resolver la Ecuación de Richards y ha sido utilizado por varios autores [2-4,14,22,26,31,33,39,41,45,48,60]. La convergencia del método de *Newton* para la Ecuación de Richards ha sido estudiada en [2–4, 41]. Este método tiene convergencia superlineal, pudiendo incluso llegar a ser cuadrática cuando las funciones no lineales involucradas tienen derivada Lipschitz continua. Sin embargo tiene algunas desventajas, el esquema usa las derivadas de las funciones y aparecen en el problema y el método solo es locamente convergente, aún cuando el uso de la solución en el tiempo anterior como solución inicial mejora considerablemente su robustez. No obstante el método podría fallar como como se muestra en un ejemplo en [4]. La convergencia del método de Newton fue estudiada en [41] para una simplificación de la Ecuación de Richards en su cuarta forma (1.11). En este trabajo se considera la conductividad hidráulica constante y no se toma en cuenta el término que corresponde a la gravedad, además es necesaria una etapa de regularización para garantizar la convergencia. El articulo utiliza un Esquema de Euler implícito para el tiempo y una discretización espacial de elementos finitos mixtos.

Se han propuesto varias estrategias para mejorar convergencia del método de Newton, por ejemplo, los métodos de continuación de Newton [61], soluciones en regiones de confianza [62]. También se desarrollaron técnicas de precondicionamiento no lineal para mejorar su rendimiento en [63, 64]. Un enfoque muy interesante es el propuesto por Brenner y Cances [48], quienes reformulan la Ecuación de Richards y utilizan un esquema de paso de tiempo adaptativo con el método de Volúmenes Finitos. Con esta estrategia logran que los sistemas no lineales satisfagan las hipótesis del Teorema de Newton-Kantoróvich consiguiendo la convergencia del método de Newton. En [38] se aplicó un esquema cuasi-Newton, más concretamente el método de Broyden. En [48] se menciona el método de Newton Semisuave como una de

las posibles alternativas al método de *Newton*, sin embargo no hemos encontrado referencias de este método aplicado a la Ecuación de Richards.

En lo que sigue usaremos las ideas propuestas en [41] para analizar la convergencia del método de *Newton*, utilizaremos los esquemas de Euler implícito y semi-implícito para el tiempo y una discretización espacial de elementos finitos de Galerkin. Usaremos la forma mixta y la cuarta forma de la Ecuación de Richards.

2.1.1 Esquema de Newton para la cuarta forma

Consideramos (1.11) con su condición inicial y estudiamos el problema solo con condiciones de Dirichlet homogéneos, pero todos los resultados expuestos en esta tesis pueden extenderse fácilmente a casos más generales.

Sea $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$ un abierto acotado con frontera $\partial \Omega$ Lipschitz continua, definimos $\Omega_T = \Omega \times (0, T)$ para una constante fija T > 0. Consideramos la siguiente ecuación diferencial parcial cuasilineal degenerada elíptica-parabólica:

$$\begin{cases} \frac{\partial b(u)}{\partial t} - \operatorname{div}(\nabla u + K(b(u))\vec{e}_z) = S & \text{sobre } \Omega_T, \\ u(x,0) = u_0(x) & \text{para todo } x \in \Omega, \\ u = 0 & \text{sobre } \partial\Omega \times (0,T). \end{cases}$$
(2.1)

Hipótesis 2.1.1. Asumimos las siguientes hipótesis en (2.1):

- (K1) $b \in C^1(\mathbb{R})$ y existe $L_b > 0$ tal que $0 \leq b'(z) \leq L_b$ para todo $z \in \mathbb{R}$.
- (K2) Existen dos constantes $C_b > 0$ y $\gamma \in (0, 1]$ tal que: $|b'(x) - b'(y)| \le C_b |x - y|^{\gamma}$ para todo $x, y \in \mathbb{R}$.
- (K3) $K \circ b \in C(\mathbb{R})$ y existe $L_K > 0$ tal que $|(K \circ b)'(z)| \leq L_K$ para todo $z \in \mathbb{R}$.
- (K4) Existe $B_K > 0$ tal que: $|K(b(z_1)) - K(b(z_2))|^2 \le B_K(b(z_1) - b(z_2))(z_1 - z_2)$ para todo $z_1, z_2 \in \mathbb{R}$.
- (K5) Existe $C_K > 0$ tal que: $|(K \circ b)'(x) - (K \circ b)'(y)| \le C_K |x - y|^{\gamma}$ para todo $x, y \in \mathbb{R}$.
- (K6) $b(u_0)$ es esencialmente acotada y $u_0 \in L^2(\Omega)$.
- (K7) $S \in L^1(\Omega \times (0,T)) \cap L^2(0,T; H^{-1}(\Omega)).$

Es importante señalar que las hipótesis (K1), (K3), (K4), (K6) y (K7) son suficientes para garantizar la existencia y unicidad de la solución débil de (2.1) [12].

Dado que b' se anula para algunos valores, el problema cambia su tipo de parabólico a elíptico. Para superar este inconveniente, empleamos un paso de regularización en (2.1). Dado $\epsilon > 0$ un parámetro de perturbación, reemplazamos la función b en (2.1) por:

$$b_{\epsilon}(z) = b(z) + \epsilon z,$$

Se satisface que $b'_{\epsilon} > \epsilon$. Para mayores detalles podemos consultar [32, 65]. Una consecuencia directa es el siguiente lema:

Lema 2.1.1. Suponiendo que se cumplen las hipótesis (K2) y (K5) entonces para todo $x, y \in \mathbb{R}$ se tiene que:

$$|b_{\varepsilon}(x) - b_{\varepsilon}(y) - b_{\varepsilon}'(y)(x - y)| \le \frac{C_b}{1 + \gamma} |x - y|^{1 + \gamma},$$
$$|(K \circ b_{\varepsilon})(x) - (K \circ b_{\varepsilon})(y) - (K \circ b_{\varepsilon})'(y)(x - y)| \le \frac{C_K}{1 + \gamma} |x - y|^{1 + \gamma}.$$

Para la demostración de este lema nos referiremos a [56, pág 350].

Problema completamente discreto

La discretización en el tiempo se logra a través de un Esquema de Euler implícito, donde denotamos el paso de tiempo por $h_t = T/N$, y $t_k = kh_t$, k = 1, 2, ..., N, para algún $N \in \mathbb{N}$ fijo y estrictamente positivo. Para el espacio utilizaremos el MEF, denotemos por \mathcal{T}_h a una descomposición regular de $\Omega \subset \mathbb{R}^d$ en *d*-simplices cerrados que denotamos como τ , tal que satisfacen las propiedades descritas en [51, p. 38] y *h* representa el diámetro de la malla. El espacio de elementos finitos de Galerkin está dado por $V_h := \{v_h \in H_0^1(\Omega)/v_{h|\tau} \in \mathcal{P}(\tau), \tau \in \mathcal{T}_h\}$, donde $\mathcal{P}(\tau)$ denota el espacio de polinomios lineales sobre cualquier símplice τ . Para obtener más detalles acerca del espacio de elementos finitos y sus propiedades, nos referimos a [51, 52]. Con la finalidad de simplificar usamos la notación $u^h(t_k) = u_k^h$. El resultado de aplicar a (2.1) un Esquema de Euler completamente implícito en el tiempo, el método de elementos finitos en el espacio y una regularización de *b* es el siguiente problema completamente discreto

En el primer paso de tiempo tomamos $u_0^h = P_h(u_0) \in V_h$, siendo $P_h : H_0^1(\Omega) \to V_h$ la proyección estándar. El esquema completamente discreto (2.2) tiene una solución única y nos referimos a [37,40] para su demostración.

Esquema de Newton

En lo que sigue daremos una formulación del método de Newton para el problema (2.2) y posteriormente enunciaremos y demostraremos el teorema principal de este apartado. Con la finalidad de simplificar la notación, omitiremos h y ϵ de las ecuaciones (es decir, $u_k^h = u_k$ y $b_{\epsilon} = b$). Usamos n para las iteraciónes del método de Newton, y k para el paso de tiempo. En consecuencia, u_k^n denota la solución en la iteración n del método de Newton y el k-ésimo paso de tiempo

$$\begin{array}{l} \left\langle \text{Dado } u_{k} \neq u_{k+1}^{n} \in V_{h}, \text{ hallar } u_{k+1}^{n+1} \in V_{h} \text{ tal que:} \\ \left\langle b'(u_{k+1}^{n})(u_{k+1}^{n+1} - u_{k+1}^{n}), v \right\rangle + \left\langle b(u_{k+1}^{n}), v \right\rangle + h_{t} \left\langle \nabla u_{k+1}^{n+1}, \nabla v \right\rangle \\ + h_{t} \left\langle (K \circ b)(u_{k+1}^{n})e_{z}, \nabla v \right\rangle \\ + h_{t} \left\langle (K \circ b)'(u_{k+1}^{n})(u_{k+1}^{n+1} - u_{k+1}^{n})e_{z}, \nabla v \right\rangle \\ = \left\langle u_{k} + h_{t}S_{k+1}, v \right\rangle \text{ para todo } v \in V_{h}. \end{array}$$

$$(2.3)$$

Observación 2.1.1. Sea $\gamma > 0$, $\{x_n\}_{n \in \mathbb{N}}$, $\{a_n\}_{n \in \mathbb{N}}$, $\{b_n\}_{n \in \mathbb{N}}$ y $\{c_n\}_{n \in \mathbb{N}}$ successores reales positivas que satisfacen:

- $a_n x_{n+1} \leq b_n x_n^{1+\gamma} + c_n x_n$ para todo $n \in \mathbb{N}$
- Existe $\lambda > 0$ tal que: $\frac{b_n x_n^{\gamma} + c_n}{a_n} \le \lambda < 1$ para todo $n \in \mathbb{N}$

entonces la sucesión $\{x_n\}_{n\in\mathbb{N}}$ converge a cero. De lo anterior si $a_n = 1$, $b_n = \beta$ y $c_n = \eta$ para todo $n \in \mathbb{N}$ es suficiente que $\beta x_0^{\gamma} + \eta < 1$ para que $\{x_n\}_{n\in\mathbb{N}}$ converja a cero.

Lema 2.1.2. Existe C > 0 independiente de los parámetros de discretización tal que:

$$\|v\|_{L^{2+2\gamma}(\Omega)}^{2+2\gamma} \le Ch^{-\gamma d} \|v\|^{2+2\gamma} \quad para \ todo \quad v \in V_h.$$

Para obtener detalles sobre esta desigualdad y su demostración podemos consultar [52, pág. 111]. Esta desigualdad fue usada en [41] para demostrar la convergencia de algunos esquemas en la ecuación de Richard.

Proposición 2.1.1. Suponga que se cumplen las hipótesis (K2), (K3), (K4), (K5) y el paso de tiempo satisface la condición $h_t \leq \frac{\varepsilon}{4L_K^2}$, entonces se verifica la siguiente desigualdad:

$$\|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{h_t}{\epsilon} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le \frac{2C(C_b + h_t \varepsilon C_K^2)}{\varepsilon^2 (1+\gamma)^2} h^{-\gamma d} \|\mathbf{e}_{k+1}^n\|^{2+2\gamma},$$

donde $\mathbf{e}_{k+1}^{n+1} = u_{k+1}^{n+1} - u_{k+1}$.

Demostración. Restando (2.3) de (2.2), se tiene:

$$\left\langle b'(u_{k+1}^n)(\mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^n), v \right\rangle + \left\langle b(u_{k+1}^n) - b(u_{k+1}), v \right\rangle + h_t \left\langle \nabla \mathbf{e}_{k+1}^{n+1}, \nabla v \right\rangle$$

$$+ h_t \left\langle \left[(K \circ b)(u_{k+1}^n) - (K \circ b)(u_{k+1}) \right] e_z, \nabla v \right\rangle$$

$$+ h_t \left\langle (K \circ b)'(u_{k+1}^n)(\mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^n) e_z, \nabla v \right\rangle = 0.$$

Escogiendo $v=\mathbf{e}_{k+1}^{n+1}$ y haciendo algunas operaciones algebraicas se deduce

$$\begin{split} \left\langle b'(u_{k+1}^{n})\mathbf{e}_{k+1}^{n+1}, \mathbf{e}_{k+1}^{n+1} \right\rangle + h_{t} \left\langle \nabla \mathbf{e}_{k+1}^{n+1}, \nabla \mathbf{e}_{k+1}^{n+1} \right\rangle \\ &\leq \left\langle b(u_{k+1}) - b(u_{k+1}^{n}) - b'(u_{k+1}^{n})(u_{k+1} - u_{k+1}^{n}), \mathbf{e}_{k+1}^{n+1} \right\rangle \\ &+ h_{t} \left\langle \left[(K \circ b)(u_{k+1}) - (K \circ b)(u_{k+1}^{n}) - (K \circ b)'(u_{k+1}^{n})(u_{k+1} - u_{k+1}^{n}) \right] e_{z}, \nabla \mathbf{e}_{k+1}^{n+1} \right\rangle \\ &+ h_{t} \left\langle \left| (K \circ b)'(u_{k+1}^{n})\mathbf{e}_{k+1}^{n+1} \right| |e_{z}|, |\nabla \mathbf{e}_{k+1}^{n+1}| \right\rangle. \end{split}$$

Usando el Lema 2.1.1, y dado que $b \geq \varepsilon$ se tiene que:

$$\begin{split} \varepsilon \|\mathbf{e}_{k+1}^{n+1}\|^2 + h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 &\leq \left\langle \frac{C_b}{1+\gamma} |u_{k+1} - u_{k+1}^n|^{1+\gamma}, |\mathbf{e}_{k+1}^{n+1}| \right\rangle \\ &+ h_t \left\langle \frac{C_K}{1+\gamma} |u_{k+1} - u_{k+1}^n|^{1+\gamma} |e_z|, |\nabla \mathbf{e}_{k+1}^{n+1}| \right\rangle \\ &+ h_t \left\langle |(K \circ b)'(u_{k+1}^n) \mathbf{e}_{k+1}^{n+1}| |e_z|, |\nabla \mathbf{e}_{k+1}^{n+1}| \right\rangle. \end{split}$$

Usando la hipótesis (K3) y de las desigualdades Young y Hölder obtenemos:

$$\varepsilon \|\mathbf{e}_{k+1}^{n+1}\|^{2} + h_{t} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^{2} \leq \frac{C_{b}^{2}}{\epsilon(1+\gamma)^{2}} \|\mathbf{e}_{k+1}^{n}\|_{L^{2+2\gamma}(\Omega)}^{2+2\gamma} + \frac{\varepsilon}{4} \|\mathbf{e}_{k+1}^{n+1}\|^{2} + \frac{h_{t}C_{K}^{2}}{(1+\gamma)^{2}} \|\mathbf{e}_{k+1}^{n}\|_{L^{2+2\gamma}(\Omega)}^{2+2\gamma} + \frac{h_{t}}{4} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^{2} + h_{t}L_{K}^{2} \|\mathbf{e}_{k+1}^{n+1}\|^{2} + \frac{h_{t}}{4} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^{2}.$$

Puesto que $h_t L_K^2 \leq \frac{\varepsilon}{4}$ y luego de algunas operaciones algebraicas se consigue

$$\frac{3\varepsilon}{4} \|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{h_t}{2} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le \frac{C_b^2 + \varepsilon h_t C_K^2}{\epsilon (1+\gamma)^2} \|\mathbf{e}_{k+1}^n\|_{L^{2+2\gamma}(\Omega)}^{2+2\gamma} + \frac{\varepsilon}{4} \|\mathbf{e}_{k+1}^{n+1}\|^2.$$

Del Lema 2.1.2 se deduce inmediatamente que

$$\|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{h_t}{\varepsilon} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le \frac{2C(C_b^2 + \varepsilon h_t C_K^2)}{\epsilon^2 (1+\gamma)^2} h^{-\gamma d} \|\mathbf{e}_{k+1}^n\|^{2+2\gamma}.$$

Teorema 2.1.1. Suponiendo las hipótesis (K1)-(K7) y que el paso de tiempo satisface la condición $h_t \leq \frac{\varepsilon}{4L_{\kappa}^2}$. Además la solución inicial u_{k+1}^0 satisface la desigualdad

$$\|u_{k+1}^{0} - u_{k+1}\| < \left[\frac{\epsilon(1+\gamma)}{\sqrt{2C(C_{b}^{2} + \varepsilon h_{t}C_{K}^{2})}}\right]^{\frac{1}{\gamma}} h^{\frac{d}{2}},$$
(2.4)

entonces el Esquema de Newton (2.3) converge con orden $1 + \gamma$.

Demostración. Usando la estimación dada en la Proposición 2.1.1 se tiene que

$$\|\mathbf{e}_{k+1}^{n+1}\| \le \frac{\sqrt{2C(C_b^2 + \varepsilon h_t C_K^2)}}{\epsilon(1+\gamma)} h^{-\frac{\gamma d}{2}} \|\mathbf{e}_{k+1}^n\|^{1+\gamma}.$$

Esto significa que si $\|\mathbf{e}_{k+1}^n\|$ converge a cero cuando *n* tiende al infinito lo hace con orden de convergencia $1 + \gamma$. Usando la Observación 2.1.1 se puede garantizar la convergencia de $\|\mathbf{e}_{k+1}^n\|$ a cero siempre que

$$\frac{\sqrt{2C(C_b^2 + \varepsilon h_t C_K^2)}}{\epsilon(1+\gamma)} h^{-\frac{\gamma d}{2}} \|\mathbf{e}_{k+1}^0\|^{\gamma} < 1,$$

lo que se cumple por la condición (2.4).

Observación 2.1.2. El Teorema 2.1.1 demuestra la convergencia local del método, es decir, la solución inicial con la que comienza el esquema de Newton, debe ser tomada lo suficientemente cerca de la solución del problema (2.2), de lo contrario el método puede fallar. Sin embargo, enfatizamos el hecho que la convergencia está asegurada solo cuando se realiza un paso de regularización. Un ejemplo donde el esquema de Newton no converge puede ser consultado en [4]. La principal ventaja de este método es el orden de convergencia $1 + \gamma$, pudiendo incluso llegar a ser cuadrática cuando las derivadas de las funciones involucradas son Lipchitz continuas.

Observación 2.1.3. Podría ser beneficioso si comenzamos las iteraciones con la solución obtenida en el paso de tiempo anterior. En este caso se puede obtener una nueva condición de convergencia para el esquema de Newton. Para este efecto primero recordamos la estimación de estabilidad probada en [31], Proposición 3.5 siguiente:

$$\|u_{k+1} - u_k\|^2 \le C \frac{h_t}{\varepsilon}.$$

Suponiendo que $h_t = O(\varepsilon^{\frac{\gamma+2}{\gamma}}) y$ del Teorema 2.1.1 se tiene que la condición de convergencia esta dada por:

$$\frac{2C(C_b^2 + \varepsilon h_t C_K^2)h_t^{\gamma}h^{-\gamma d}}{\varepsilon^{2+\gamma}} < 1.$$

 \square

Observación 2.1.4. Según la condición (2.4), la convergencia del método de Newton depende del tamaño de la malla espacial h, es decir, que al hacer la malla espacial más fina inevitablemente se llega a una divergencia del método. Este resultado teórico está apoyado por el ejemplo numérico presentado en [4] y los experimentos numéricos presentados en el capítulo 3 de esta tesis, donde al hacer cada vez más pequeño el tamaño de la malla h en algún momento el método falla.

2.1.2 Esquema de Newton para la forma mixta

Consideramos la forma mixta de la Ecuación de Richards (1.8) con su condición inicial y condiciones de borde de Dirichlet homogéneas. Sea $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$ un abierto acotado con frontera $\partial \Omega$ Lipschitz continua, definimos $\Omega_T = \Omega \times (0, T)$ para una constante fija T > 0. Contemplamos la ecuación diferencial parcial cuasi lineal degenerada elíptica-parabólica siguiente:

$$\begin{cases} \frac{\partial \theta(\Psi)}{\partial t} - \operatorname{div}(K(\Psi)(\nabla \Psi + \vec{e}_z) = S & \text{sobre } \Omega_T, \\ \Psi(x, 0) = \Psi_0(x) & \text{para todo } x \in \Omega, \\ \Psi = 0 & \text{sobre } \partial\Omega \times (0, T). \end{cases}$$
(2.5)

Hipótesis 2.1.2. Asumimos las siguientes hipótesis en (2.5) [15]:

(M1) $\theta \in C^1(\mathbb{R})$ y existe $L_{\theta} > 0$ tal que $0 \leq \theta'(z) \leq L_{\theta}$ para todo $z \in \mathbb{R}$.

- (M2) Existen dos constantes $C_{\theta} > 0$ $y \gamma \in (0, 1]$ tal que: $|\theta'(x) - \theta'(y)| \le C_{\theta}|x - y|^{\gamma}$ para todo $x, y \in \mathbb{R}$.
- (M3) $K \in C(\mathbb{R})$ y existen $K_M, K_m > 0$ tal que $K_m \leq K(z) \leq K_M$ para todo $z \in \mathbb{R}$.
- (M4) Existe $C_K > 0$ tal que: $|K(\theta(z_1)) - K(\theta(z_2))|^2 \le C_K(\theta(z_1) - \theta(z_2))(z_1 - z_2)$ para todo $z_1, z_2 \in \mathbb{R}$.
- (M5) $\Psi_0 \in L^2(\Omega)$.

 $(M6) \ S \in L^1(\Omega \times (0,T)) \cap L^2(0,T;H^{-1}(\Omega)).$

Las hipótesis (M1), (M3), (M4), (M5) y (M6) son suficientes para garantizar la existencia y unicidad de la solución débil de (2.5) [12].

Es importante resaltar que no fue posible una demostración formal de la convergencia del Método de *Newton* para la forma mixta de la Ecuación de Richards (2.5) cuando se utiliza un Esquema de Euler totalmente implícito, sin embargo los experimentos numéricos presentados en [4] sugieren que en este caso una condición similar a (2.4) podría cumplirse. Cuando usamos un Esquema de Euler explícito en el termino de conductividad hidráulica e implícito en todos los demás términos si podemos conseguir una demostración, usando las mismas ideas del teorema anterior. Este Esquema de Euler ha sido usado por otros autores [15, 32]. A continuación presentamos una formulación completamente discreta de (2.5) utilizando el Esquema de Euler semi-implícito antes mencionado y un esquema Elementos Finitos de Galerkin para el espacio. Esta formulación puede ser consultada en [15].

 $\begin{cases} \text{Dado } \Psi_k^h \in V_h, \text{ hallar } \Psi_{k+1}^h \in V_h \text{ tal que:} \\ \left\langle \theta(\Psi_{k+1}^h), v_h \right\rangle + h_t \left\langle K(\Psi_k^h)(\nabla \Psi_{k+1}^h + \vec{e_z}), \nabla v_h \right\rangle = \left\langle \theta(\Psi_k^h) + h_t S_{k+1}, v_h \right\rangle \qquad (2.6) \\ \text{para cada } v_h \in V_h. \end{cases}$

Para el primer paso de tiempo tomamos $\Psi_0^h = P_h(\Psi_0) \in V_h$.

Por las hipótesis (M1)-(M6), el lado izquierdo de (2.6) es un operador maximal monótono y coercivo de $H_0^1(\Omega)$ en $H^{-1}(\Omega)$. El lado derecho de (2.6) es un funcional lineal acotado en $H_0^1(\Omega)$. Por lo tanto, según el Corolario 2.4 de [66], el problema (2.6) tiene una única solución débil $\Psi_{k+1} \in H_0^1(\Omega)$.

Con la finalidad de simplificar la notación, omitiremos el superíndice h de las ecuaciones $\Psi_k^h = \Psi_k$. Usamos n para las iteraciónes del método de Newton, y k para el paso de tiempo. El método de Newton para el problema no lineal (2.6) está dado por:

$$\begin{cases} \text{Dado } \Psi_k, \Psi_{k+1}^n \in V_h, \text{ hallar } \Psi_{k+1}^{n+1} \in V_h \text{ tal que:} \\ \left\langle \theta'(\Psi_{k+1}^n)(\Psi_{k+1}^{n+1} - \Psi_{k+1}^n), v \right\rangle + h_t \left\langle K(\Psi_k)(\nabla \Psi_{k+1}^{n+1} + \vec{e_z}), \nabla v \right\rangle \\ = \left\langle \theta_k - \theta_{k+1}^n + h_t S_{k+1}, v \right\rangle \text{ para todo } v \in V_h. \end{cases}$$

$$(2.7)$$

En el esquema anterior y en todo lo que se sigue de la tesis usaremos la notación $\Psi^n(t_{k+1}) = \Psi^n_{k+1}$ y $\theta(\Psi^n_{k+1}) = \theta^n_{k+1}$.

Proposición 2.1.2. Si se satisfacen las hipótesis (M1), (M2) y (M3), entonces se verifica la siguiente desigualdad:

$$\|e_{k+1}^{n+1}\|^2 + 2C_p \|\nabla e_{k+1}^{n+1}\|^2 \le \frac{4CC_{\theta}^2 C_p^4}{h_t^2 K_m^2 (1+\gamma)^2} h^{-\gamma d} \|e_{k+1}^n\|^{2+2\gamma},$$

donde $e_{k+1}^{n+1} = \Psi_{k+1}^{n+1} - \Psi_{k+1}$, C_p es la constante de la desigualdad de Poincaré y C es la constante del Lema 2.1.2.

Demostración. Restando (2.7) de (2.6), se tiene:

$$\left\langle \theta'(\Psi_{k+1}^n)(e_{k+1}^{n+1}-e_{k+1}^n),v\right\rangle + \left\langle \theta(\Psi_{k+1}^n) - \theta(\Psi_{k+1}),v\right\rangle + h_t \left\langle K(\Psi_k)\nabla e_{k+1}^{n+1},\nabla v\right\rangle = 0.$$

Escogiendo $v=e_{k+1}^{n+1}$ y luego de algunas operaciones algebraicas se deduce

$$\left\langle \theta'(\Psi_{k+1}^{n})e_{k+1}^{n+1}, e_{k+1}^{n+1} \right\rangle + h_t \left\langle K(\Psi_k) \nabla e_{k+1}^{n+1}, \nabla e_{k+1}^{n+1} \right\rangle \\ \leq \left\langle \theta(\Psi_{k+1}) - \theta(\Psi_{k+1}^{n}) - \theta'(\Psi_{k+1}^{n})(\Psi_{k+1} - \Psi_{k+1}^{n}), e_{k+1}^{n+1} \right\rangle.$$

Usando la hipótesis (M1), (M2), (M3) y el Lema 2.1.1 se sigue

$$h_t K_m \|\nabla e_{k+1}^{n+1}\|^2 \le \left\langle \frac{C_\theta}{1+\gamma} |\Psi_{k+1} - \Psi_{k+1}^n|^{1+\gamma}, |\nabla e_{k+1}^{n+1}| \right\rangle.$$

Usando las desigualdades de Poincaré, Young y Hölder nosotros obtenemos

$$\frac{h_t K_m}{2C_p^2} \|e_{k+1}^{n+1}\|^2 + \frac{h_t K_m}{2} \|\nabla e_{k+1}^{n+1}\|^2 \le \frac{C_\theta^2 C_p^2}{h_t K_m (1+\gamma)^2} \|e_{k+1}^n\|_{L^{2+2\gamma}(\Omega)}^{2+2\gamma} + \frac{h_t K_m}{4C_p^2} \|e_{k+1}^{n+1}\|^2.$$

Del Lema 2.1.2 se deduce inmediatamente que

$$\frac{h_t K_m}{4C_p^2} \|e_{k+1}^{n+1}\|^2 + \frac{h_t K_m}{2} \|\nabla e_{k+1}^{n+1}\|^2 \le C \frac{C_\theta^2 C_p^2}{h_t K_m (1+\gamma)^2} h^{-\gamma d} \|e_{k+1}^{n+1}\|^{2+2\gamma}$$

Simplificando se tiene inmediatamente el resultado.

Teorema 2.1.2. Suponiendo las hipótesis (M1)-(M6) y que la solución inicial con que comienza el método Ψ_{k+1}^0 satisface

$$\|\Psi_{k+1}^{0} - \Psi_{k+1}\| < \left[\frac{K_m(1+\gamma)}{2\sqrt{C}C_{\theta}C_p^2}\right]^{\frac{1}{\gamma}} h_t^{\frac{1}{\gamma}} h^{\frac{d}{2}},$$
(2.8)

entonces el esquema de Newton (2.7) converge con orden $1 + \gamma$.

Demostración. Usando la estimación dada en la Proposición 2.1.2 se tiene que

$$\|e_{k+1}^{n+1}\| \le \frac{2\sqrt{C}C_{\theta}C_{p}^{2}}{h_{t}K_{m}(1+\gamma)}h^{-\frac{\gamma d}{2}}\|e_{k+1}^{n}\|^{1+\gamma}.$$

Usando la Observación 2.1.1 se puede garantizar la convergencia de $\|e_{k+1}^n\|$ a cero siempre que

$$\frac{2\sqrt{CC_{\theta}C_{p}^{2}}}{h_{t}K_{m}(1+\gamma)}h^{-\frac{\gamma d}{2}}\|e_{k+1}^{0}\|^{\gamma} < 1,$$

lo que se cumple por la condición (2.8).
Observación 2.1.5. En este caso se puede obtener la demostración sin necesidad de usar una regularización en la función θ , lo que significa una gran ventaja en comparación con el esquema (2.3), además que al resolver la forma mixta se obtiene la cabeza de presión Ψ , que es la variable de interés en el modelo.

Observación 2.1.6. El Teorema 2.1.2 garantiza que si el método converge, lo hace con orden $1 + \gamma$ en $L^2(\Omega)$, teniendo convergencia cuadrática cuando θ' es Lipschitz continua, pero el método es únicamente localmente convergente. En el capítulo 3 de esta tesis damos un ejemplo donde este esquema no es convergente. De la Proposición 2.1.2 se deduce que si el método converge en $L^2(\Omega)$, lo hace también en $H^1(\Omega)$.

Observación 2.1.7. La convergencia del método de Newton depende del tamaño de la malla temporal h_t y espacial h, pues en la condición (2.8) uno puede tomar una solución inicial fija cualquiera y hacer la malla espacial más fina o el paso de tiempo más pequeño, lo que según (2.8) conduce inevitablemente a una divergencia del método.

Observación 2.1.8. El método de Newton para la forma mixta usando un Esquema de Euler completamente implícito puede ser consultado en [4], aunque sin ninguna demostración de su convergencia. Sin embargo los autores presentan ejemplos donde se pone en evidencia que la convergencia del método también depende del tamaño de la malla espacial.

2.2 Método de Picard Modificado

Este esquema fue propuesto por Celia et. al. en 1990 [3] para la forma mixta de la Ecuación de Richards, y posteriormente fue usado por varios autores [4, 14, 16, 19, 23, 38, 46, 47, 60]. El esquema coincide con el método de *Newton* en el caso de conductividad hidráulica constante. El método de *Picard Modificado* es solo linealmente convergente, pero según Lehmann et al. en 1998 [14] es más robusto que el método de *Newton*, llegando a esta conclusión mediante algunos experimentos numéricos. La idea del método es discretizar la no linealidad que involucra a θ cuadráticamente, mientras que la no linealidad de K se aproxima linealmente, por lo tanto el método solo es linealmente convergente, ademas el método aún involucra el cálculo de derivadas. Es importante señalar que la convergencia de este esquema también podría fallar como se demuestra en uno de los ejemplos propuestos en [4]. Para construir el esquema de *Picard Modificado* [3], comenzamos usando un esquema de linealización de *Picard* estándar [2] en la Ecuación de Richards en su forma mixta (2.5).

$$\begin{cases} \text{Dado } \Psi_k, \Psi_k^n \in V_h, \text{ hallar } \Psi_{k+1}^{n+1} \in V_h \text{ tal que:} \\ \left\langle \theta_{k+1}^{n+1}, v \right\rangle + h_t \left\langle K_{k+1}^n (\nabla \Psi_{k+1}^{n+1} + \vec{e_z}), v \right\rangle = \left\langle \theta_k + h_t S_{k+1}, v \right\rangle \end{aligned}$$
(2.9)
para todo $v \in V_h.$

La clave del método es expandir θ en una serie de Taylor truncada en una vecindad de Ψ_{k+1}^n y evaluar la expansión en Ψ_{k+1}^{n+1}

$$\theta_{k+1}^{n+1} \approx \theta_{k+1}^n + \frac{d\theta_{k+1}^n}{d\Psi} (\Psi_{k+1}^{n+1} - \Psi_{k+1}^n) = \theta_{k+1}^n + C_{k+1}^n (\Psi_{k+1}^{n+1} - \Psi_{k+1}^n)$$

Reemplazando (2.2) en (2.9) se tiene finalmente el esquema de *Picard Modificado* llamado esquema de *Picard*. Este método es el siguiente:

$$\begin{cases} \text{Dado } \Psi_{k}, \Psi_{k}^{n} \in V_{h}, \text{ hallar } \Psi_{k+1}^{n+1} \in V_{h} \text{ tal que:} \\ \left\langle C_{k+1}^{n}(\Psi_{k+1}^{n+1} - \Psi_{k+1}^{n}), v \right\rangle + h_{t} \left\langle K_{k+1}^{n}(\nabla \Psi_{k+1}^{n+1} + e_{\overline{z}}), v \right\rangle \\ = \left\langle \theta_{k} - \theta_{k+1}^{n} + h_{t}S_{k+1}, v \right\rangle \\ \text{para todo } v \in V_{h}. \end{cases}$$
(2.10)

Este esquema puede ser consultado en [3,4], donde se proponen diversos experimentos numéricos para estudiar su convergencia, encontrando que existe un aumento cuantitativo de la robustez del método si lo comparamos con el método de *Newton*. Si el método de *Picard* falla, también el de *Newton*, pero si el método de *Newton* falla, no necesariamente falla el método de *Picard*. Un ejemplo de esto se puede consultar en [4], donde el método de *Newton* no converge, sin embargo el método de *Picard* sí lo hace. Pero al aumentar el número de elementos de la malla, el método de *Picard* falla. De este ejemplo también se desprende que la convergencia aún depende del tamaño de la malla espacial. Es importante resaltar que aún cuando existe un aumento de la robustez, el método de *Picard* sigue teniendo convergencia local.

2.2.1 Esquema de Picard para la cuarta forma

La convergencia del método de *Picard Modificado* se estudia en [41], para la cuarta forma (2.1) usando un Esquema de Euler totalmente implícito en el tiempo y el método de elementos finitos mixtos para el espacio. En lo que sigue analizaremos la convergencia del método de *Picard*. Asumimos las hipótesis (K1)-(K4), (K6) y (K7) en (2.2). Usamos un Esquema de Euler totalmente implícito y el método de elementos finitos de Galerkin para el espacio. Al igual que para el método de *Newton* es necesaria una regularización de la función b. El esquema de *Picard Modificado* para el problema totalmente discreto y regularizado (2.2) es el siguiente

$$\begin{cases} \text{Dado } u_k \neq u_{k+1}^n \in V_h, \text{ hallar } u_{k+1}^{n+1} \in V_h \text{ tal que:} \\ \left\langle b'(u_{k+1}^n)(u_{k+1}^{n+1} - u_{k+1}^n), v \right\rangle + \left\langle b(u_{k+1}^n), v \right\rangle + h_t \left\langle \nabla u_{k+1}^{n+1}, \nabla v \right\rangle \\ + h_t \left\langle (K \circ b)(u_{k+1}^n)e_z, \nabla v \right\rangle \\ = \left\langle u_k + h_t S_{k+1}, v \right\rangle \text{ para todo } v \in V_h. \end{cases}$$

$$(2.11)$$

Proposición 2.2.1. Si se satisfacen las hipótesis (K2) y (K3), entonces se verifica la siguiente desigualdad:

$$\|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{h_t}{\epsilon} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le \frac{CC_b^2}{\varepsilon^2 (1+\gamma)^2} h^{-\gamma d} \|\mathbf{e}_{k+1}^n\|^{2+2\gamma} + \frac{h_t L_K^2}{\varepsilon} \|\mathbf{e}_{k+1}^n\|^2.$$

Demostración. Restando (2.11) de (2.2), se tiene:

$$\left\langle b'(u_{k+1}^n)(\mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^n), v \right\rangle + \left\langle b(u_{k+1}^n) - b(u_{k+1}), v \right\rangle + h_t \left\langle \nabla \mathbf{e}_{k+1}^{n+1}, \nabla v \right\rangle$$
$$+ h_t \left\langle \left[(K \circ b)(u_{k+1}^n) - (K \circ b)(u_{k+1}) \right] e_z, \nabla v \right\rangle = 0.$$

Escogiendo $v = \mathbf{e}_{k+1}^{n+1}$, se deduce

$$\left\langle b'(u_{k+1}^{n})\mathbf{e}_{k+1}^{n+1}, \mathbf{e}_{k+1}^{n+1} \right\rangle + h_t \left\langle \nabla \mathbf{e}_{k+1}^{n+1}, \nabla \mathbf{e}_{k+1}^{n+1} \right\rangle$$

$$\leq \left\langle b(u_{k+1}) - b(u_{k+1}^{n}) - b'(u_{k+1}^{n})(u_{k+1} - u_{k+1}^{n}), \mathbf{e}_{k+1}^{n+1} \right\rangle$$

$$+ h_t \left\langle \left[(K \circ b)(u_{k+1}) - (K \circ b)(u_{k+1}^{n}) \right] e_z, \nabla \mathbf{e}_{k+1}^{n+1} \right\rangle.$$

Dado que $b \ge \varepsilon$ y usando la primera desigualdad del Lema 2.1.1 se tiene:

$$\varepsilon \|\mathbf{e}_{k+1}^{n+1}\|^2 + h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \leq \left\langle \frac{C_b}{1+\gamma} |u_{k+1} - u_{k+1}^n|^{1+\gamma}, |\mathbf{e}_{k+1}^{n+1}| \right\rangle \\ + h_t \left\langle \left[(K \circ b)(u_{k+1}) - (K \circ b)(u_{k+1}^n) \right] e_z, \nabla \mathbf{e}_{k+1}^{n+1} \right\rangle.$$

Usando las desigualdades Young y Hölder, obtenemos:

$$\varepsilon \|\mathbf{e}_{k+1}^{n+1}\|^2 + h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le \frac{C_b^2}{2\epsilon(1+\gamma)^2} \|\mathbf{e}_{k+1}^n\|_{L^{2+2\gamma}(\Omega)}^{2+2\gamma} + \frac{\varepsilon}{2} \|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{h_t}{2} \|(K \circ b)(u_{k+1}) - (K \circ b)(u_{k+1}^n)\|^2 + \frac{h_t}{2} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2.$$

De la hipótesis (K3)se deduce que $K \circ b$ es Lipschitz continua y haciendo algunas operaciones algebraicas obtenemos

$$\|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{h_t}{\varepsilon} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le \frac{C_b^2}{\epsilon^2 (1+\gamma)^2} \|\mathbf{e}_{k+1}^n\|_{L^{2+2\gamma}(\Omega)}^{2+2\gamma} + \frac{h_t L_K^2}{\varepsilon} \|\mathbf{e}_{k+1}^n\|^2.$$

Del Lema 2.1.2 se deduce inmediatamente que:

$$\|\mathbf{e}_{k+1}^{n+1}\|^{2} + \frac{h_{t}}{\varepsilon} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^{2} \leq \frac{CC_{b}^{2}}{\epsilon^{2}(1+\gamma)^{2}} h^{-\gamma d} \|\mathbf{e}_{k+1}^{n}\|^{2+2\gamma} + \frac{h_{t}L_{K}^{2}}{\varepsilon} \|\mathbf{e}_{k+1}^{n}\|^{2}.$$

Teorema 2.2.1. Suponiendo que se cumplen las hipótesis (K2) y (K3), el paso de tiempo satisface la condición $h_t \leq \frac{\varepsilon}{L_K^2}$ y que la solución con que comienza el método u_{k+1}^0 satisface la desigualdad:

$$\|u_{k+1}^0 - u_{k+1}\| < \left[\frac{\sqrt{\varepsilon - h_t L_K^2} (1+\gamma)\sqrt{\epsilon}}{\sqrt{C}C_b}\right]^{\frac{1}{\gamma}} h^{\frac{d}{2}}, \qquad (2.12)$$

entonces el esquema de Picard Modificado (2.11) converge con orden de convergencia al menos lineal.

Demostración. Usando la estimación dada en la Proposición 2.2.1 se tiene que

$$\|\mathbf{e}_{k+1}^{n+1}\|^{2} \leq \frac{CC_{b}^{2}}{\epsilon^{2}(1+\gamma)^{2}}h^{-\gamma d}\|\mathbf{e}_{k+1}^{n}\|^{2+2\gamma} + \frac{h_{t}L_{K}^{2}}{\varepsilon}\|\mathbf{e}_{k+1}^{n}\|^{2}$$

De esta desigualdad se deduce que si $\|\mathbf{e}_{k+1}^n\|$ converge a cero cuando *n* tiende al infinito, lo hace con orden de convergencia por lo menos lineal. Usando la Observación 2.1.1 se puede garantizar la convergencia de $\|\mathbf{e}_{k+1}^n\|$ a cero siempre que

$$\frac{CC_b^2}{\epsilon^2 (1+\gamma)^2} h^{-\gamma d} \|\mathbf{e}_{k+1}^0\|^{2\gamma} + \frac{h_t L_K^2}{\varepsilon} < 1,$$

Esto se cumple por (2.12) y la desigualdad $h_t \leq \frac{\varepsilon}{L_W^2}$.

Observación 2.2.1. La desigualdad dada en la Proposición 2.2.1, demuestra que si el método de Picard converge, no solo lo hace en $L^2(\Omega)$, sino también en el espacio $H^1(\Omega)$. La desigualdad probada en el Teorema 2.2.1 muestra que la convergencia del método aún depende del tamaño de la malla espacial, pues si hacemos cada vez más pequeño h en algún momento el Método de Picard fallará. Además es importante destacar que para obtener la prueba es imprescindible usar una regularización sobre la función b.

Observación 2.2.2. Note que para la demostración del método de Picard fue necesaria la condición $h_t \leq \frac{\varepsilon}{L_K^2}$, que es una hipótesis más relajada que la condición usada en prueba de la convergencia del esquema de Newton $h_t \leq \frac{\varepsilon}{4L_K^2}$. Si usamos esta condición en el método de Picard es fácil verificar que

$$\frac{\epsilon(1+\gamma)}{\sqrt{2C(C_b^2+\varepsilon h_t C_K^2)}} < \frac{\sqrt{\varepsilon - h_t L_K^2} (1+\gamma)\sqrt{\epsilon}}{\sqrt{C}C_b}.$$

Esta desigualdad nos permite comparar (2.4) y (2.12), de donde se deduce inmediatamente que la convergencia del método de Newton implica la convergencia del método de Picard.

2.3 Método L-scheme

El *L-scheme* fue propuesto por Slodicka [15], luego fue estudiado en [31] para elementos finitos mixtos, posteriormente en [4, 49] para el método de elementos finitos Galekin. No se han encontrado reportes de este esquema de linealización con el método de volúmenes finitos, aunque la teoría desarrollada en varios de estos trabajos y también en esta tesis sugiere que el método debería funcionar con cualquier método escogido para discretizar el espacio. Este es el único esquema que aprovecha la monotonicidad de la función involucrada en la derivada temporal y se caracteriza por ser un esquema robusto, aunque solo linealmente convergente, no involucra ningún cálculo de derivadas y además la convergencia no depende del tamaño de la malla espacial aunque en algunas ocasiones es necesaria alguna restricción en el paso de tiempo.

En [4] se reportó que los sistemas lineales de ecuaciones algebraicas que se obtienen con el *L-scheme* están mucho mejor condicionados que los sistemas obtenidos usando los métodos de *Newton* y *Picard*, debido a esto existen ejemplos donde el *L-scheme* puede ser incluso más rápido en tiempo de máquina que el método *Newton* aún cuando solo tiene convergencia lineal. Este fenómeno se puede observar en algunos ejemplos presentados en [4]. Restringiremos el análisis a condiciones de borde de Dirichlet homogéneas solo por simplicidad, aún cuando la extensión a condiciones más generales suelen ser simples.

2.3.1 L-esquema para la cuarta forma

El *L-scheme* para (2.1) fue estudiado en [31] para una discretización espacial de elementos finitos mixtos y dos esquemas de Euler: el uno totalmente implícito, donde es necesario poner una pequeña restricción sobre el paso de tiempo y el otro es un Esquema de Euler explicito en la función de conductividad hidráulica e implícito en los demás términos, este no necesita que se imponga ninguna condición sobre el paso de tiempo. Es importante destacar que para probar la convergencia del *L-scheme* no es necesario usar ningún tipo de regularización, lo que significa una gran ventaja sobre otros esquemas. Se asume que se cumplen las hipótesis (K1), (K3), (K4), (K6) y (K7) en (2.1).

En lo que sigue demostraremos los teoremas relacionados con la convergencia del *L-scheme* para una discretización espacial de elementos finitos de Galerkin, aunque las pruebas de los teoremas presentadas en esta tesis difieren un poco de las presentadas en [31]. El problema completamente discreto de (2.1) sin regularizar la función *b* queda como se sigue:

$$\begin{cases} \text{Dado } u_k \in V_h, \text{ hallar } u_{k+1} \in V_h \text{ tal que:} \\ \langle b(u_{k+1}), v \rangle + h_t \langle \nabla u_{k+1} + K(b(u_{k+1})) \vec{e_z}, \nabla v \rangle = \langle u_k^h + h_t S_{k+1}, v \rangle \\ \text{para todo } v \in V_h. \end{cases}$$
(2.13)

El método de linealización L-scheme para el problema (2.13) es el siguiente y puede ser consultado en [31].

$$\begin{cases} \text{Dado } u_k \neq u_{k+1}^n \in V_h, \text{ hallar } u_{k+1}^{n+1} \in V_h \text{ tal que:} \\ L \left\langle u_{k+1}^{n+1} - u_{k+1}^n, v \right\rangle + \left\langle b(u_{k+1}^n), v \right\rangle + h_t \left\langle \nabla u_{k+1}^{n+1}, \nabla v \right\rangle \\ + h_t \left\langle (K \circ b)(u_{k+1}^n) \vec{e_z}, \nabla v \right\rangle \\ = \left\langle u_k + h_t S_{k+1}, v \right\rangle \text{ para todo } v \in V_h, \end{cases}$$

$$(2.14)$$

donde L es una constante positiva relacionada con la la función b.

Lema 2.3.1. Asumiendo la hipótesis (K1), se tiene que para todo $v_1, v_2 \in L^2(\Omega)$ se verifica

$$||b(v_1) - b(v_2)||^2 \le L_b \langle b(v_1) - b(v_2), v_1 - v_2 \rangle.$$

Demostración. De la hipótesis (K1) se deduce que b es Lipschitz continua, luego se tiene

$$|b(x) - b(y)| \le L_b |x - y|$$
 para todo $x, y \in \mathbb{R}$.

Multiplicando los dos lados por |b(x) - b(y)| se sigue

$$|b(x) - b(y)|^2 \le L_b |x - y| |b(x) - b(y)|$$
 para todo $x, y \in \mathbb{R}$.

De (K1) se deduce que b es monótona creciente por lo que $(x - y)(b(x) - b(y)) \ge 0$ para todo $x, y \in \mathbb{R}$ obteniendo

$$|b(x) - b(y)|^2 \le L_b(x - y)(b(x) - b(y))$$
 para todo $x, y \in \mathbb{R}$

Tomando $v_1, v_2 \in L^2(\Omega)$, remplazando en la ecuación anterior e integrando sobre Ω se concluye el resultado.

Lema 2.3.2. Suponiendo (K4), se tiene que para todo $v_1, v_2 \in L^2(\Omega)$ se verifica la siguiente desigualdad:

$$||K(b(v_1)) - K(b(v_2))||^2 \le L_K \langle b(v_1) - b(v_2), v_1 - v_2 \rangle.$$

Demostración. La demostración se sigue inmediatamente de la hipótesis (K4)

Teorema 2.3.1. Asumiendo que se cumplen las hipótesis (K1) y (K4), $L > L_b$ y el paso de tiempo satisface la condición $h_t L_k \leq 1$, entonces el esquema (2.14) converge a la solución de (2.13), para cualquier solución inicial que utilicemos para comenzar las iteraciones el método, el orden de convergencia es lineal y su tasa de convergencia α es

$$\alpha = \sqrt{\frac{1}{1 + \frac{h_t}{LC_p^2}}}.$$

Demostración. Restando (2.14) de (2.13), se tiene:

$$L \left\langle \mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^{n}, v \right\rangle + \left\langle b(u_{k+1}^{n}) - b(u_{k+1}), v \right\rangle + h_t \left\langle \nabla \mathbf{e}_{k+1}^{n+1}, \nabla v \right\rangle \\ + h_t \left\langle \left[(K \circ b)(u_{k+1}^{n}) - (K \circ b)(u_{k+1}) \right] \vec{e_z}, \nabla v \right\rangle = 0.$$

Escogiendo $v = \mathbf{e}_{k+1}^{n+1}$ y luego de algunas operaciones algebraicas se deduce

$$\frac{L}{2} \|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{L}{2} \|\mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^n\|^2 + h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2
\leq \frac{L}{2} \|\mathbf{e}_{k+1}^n\|^2 + \langle b(u_{k+1}) - b(u_{k+1}^n), \mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^n \rangle
- \langle b(u_{k+1}^n) - b(u_{k+1}), u_{k+1}^n - u_{k+1} \rangle
+ h_t \langle \left[(K \circ b)(u_{k+1}^n) - (K \circ b)(u_{k+1}) \right] \vec{e_z}, \nabla \mathbf{e}_{k+1}^{n+1} \rangle.$$

Usando las desigualdades Young y Hölder obtenemos

-

$$\frac{L}{2} \|\mathbf{e}_{k+1}^{n+1}\|^{2} + \frac{L}{2} \|\mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^{n}\|^{2} + h_{t} \|\nabla\mathbf{e}_{k+1}^{n+1}\|^{2}
\leq \frac{L}{2} \|\mathbf{e}_{k+1}^{n}\|^{2} + \frac{1}{2L} \|b(u_{k+1}) - b(u_{k+1}^{n})\|^{2} + \frac{L}{2} \|\mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^{n}\|^{2}
- \langle b(u_{k+1}^{n}) - b(u_{k+1}), u_{k+1}^{n} - u_{k+1} \rangle
+ \frac{h_{t}}{2} \|(K(b(u_{k+1}^{n})) - K(b(u_{k+1}))\|^{2} + \frac{h_{t}}{2} \|\nabla\mathbf{e}_{k+1}^{n+1}\|^{2}.$$

Luego de algunas simplificaciones se consigue

$$L \|\mathbf{e}_{k+1}^{n+1}\|^{2} + h_{t} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^{2} \leq L \|\mathbf{e}_{k+1}^{n}\|^{2} + \frac{1}{L} \|b(u_{k+1}) - b(u_{k+1}^{n})\|^{2} -2 \left\langle b(u_{k+1}^{n}) - b(u_{k+1}), u_{k+1}^{n} - u_{k+1} \right\rangle + h_{t} \|(K(b(u_{k+1}^{n})) - K(b(u_{k+1}))\|^{2}.$$

Usando los Lemas 2.3.1 y 2.3.2 en el lado derecho se sigue

$$L \|\mathbf{e}_{k+1}^{n+1}\|^2 + h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le L \|\mathbf{e}_{k+1}^n\|^2 + \left[\frac{L_b}{L} - 2 + h_t C_K\right] \left\langle b(u_{k+1}^n) - b(u_{k+1}), u_{k+1}^n - u_{k+1} \right\rangle.$$

Puesto que $L > L_b$ y $h_t C_K < 1$, el último término de la derecha es negativo por lo tanto

$$L \|\mathbf{e}_{k+1}^{n+1}\|^2 + h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le L \|\mathbf{e}_{k+1}^n\|^2.$$

De la desigualdad anterior se concluye las dos desigualdades siguientes:

$$\|\mathbf{e}_{k+1}^{n+1}\| \le \alpha \|\mathbf{e}_{k+1}^{n}\|,\tag{2.15}$$

$$\|\nabla \mathbf{e}_{k+1}^{n+1}\| \le \sqrt{\frac{L}{h_t}} \|\mathbf{e}_{k+1}^n\|.$$
(2.16)

La desigualdad (2.15) se obtiene aplicando la desigualdad de Poincaré en el segundo término del lado izquierdo. De (2.15) se concluye directamente la convergencia global del esquema con un orden por lo menos lineal.

Observación 2.3.1. La desigualdad (2.15) prueba que el esquema (2.14) converge linealmente con una tasa convergencia α en el espacio $L^2(\Omega)$ y la desigualdad (2.16) demuestra que es convergente también en el espacio $H^1(\Omega)$. La convergencia de (2.14) es global porque es independiente de la solución inicial u_{k+1}^0 escogida para iniciar el método. Sin embargo, será más beneficioso si uno comienza las iteraciones con la solución obtenida en el paso de tiempo anterior.

Observación 2.3.2. En el Teorema 2.3.1 hemos tomado las hipótesis que permiten garantizar la convergencia del esquema: $L > L_b y h_t C_K < 1$ que son las que aparecen en [31]. Sin embargo debemos notar que es suficiente imponer la siguiente relación entre L y el paso de tiempo h_t

$$h_t \le \frac{2L - L_b}{LC_K}.$$

En esta desigualdad se puede apreciar que basta tomar $L > \frac{L_b}{2}$. En cualquiera de los dos casos, son restricciones relativamente leves, ya que no dependen del tamaño de la malla espacial y peor aún de algún parámetro de regularización.

Observación 2.3.3. La tasa de convergencia α depende del paso de tiempo h_t y el parámetro L, pero es independiente del tamaño de la malla espacial h. Pasos de tiempo más grandes y valores de L más pequeños dan como resultado tasas de convergencia más pequeñas y por lo tanto, el método debería converger más rápido.

A continuación presentamos el *L-scheme* para la cuarta forma de la Ecuación de Richards (2.1), pero esta vez usamos un Esquema de Euler semi-implícito: explícito en el término de conductividad hidráulica K e implícito en los demás términos. Este esquema fue usado en [31] con elementos finitos mixtos. La formulación completamente discreta de (2.1) con un Esquema de Euler semi-implícito, usando elementos finitos de Galerkin para el espacio y sin ningún tipo de regularización es como se sigue:

$$\begin{cases} \text{Dado } u_k \in V_h, \text{ hallar } u_{k+1} \in V_h \text{ tal que:} \\ \langle b(u_{k+1}), v \rangle + h_t \langle \nabla u_{k+1} + K(b(u_k)) \vec{e_z}, \nabla v \rangle = \langle u_k + h_t S_{k+1}, v \rangle \\ \text{para cada } v \in V_h. \end{cases}$$
(2.17)

El L-scheme para el problema (2.17) es el siguiente.

$$\begin{cases}
\text{Dado } u_k \neq u_{k+1}^n \in V_h \text{ hallar } u_{k+1}^{n+1} \in V_h \text{ tal que:} \\
L \left\langle u_{k+1}^{n+1} - u_{k+1}^n, v \right\rangle + \left\langle b(u_{k+1}^n), v \right\rangle + h_t \left\langle \nabla u_{k+1}^{n+1} + K(b(u_k)\vec{e_z}, \nabla v) \right\rangle \\
= \left\langle u_k + h_t S_{k+1}, v \right\rangle \text{ para todo } v \in V_h,
\end{cases}$$
(2.18)

donde L > 0 es una constante. Como probaremos en el teorema siguiente el esquema (2.18) converge sin que sea necesario imponer ninguna restricción en el tiempo.

Teorema 2.3.2. Suponiendo que se cumplen las hipótesis (K1) y (K4) y si $2L > L_b$, entonces el esquema (2.18) converge para cualquier solución inicial con que se comience las iteraciones del método, con un orden de convergencia lineal y tasa de convergencia α

$$\alpha = \sqrt{\frac{1}{1 + \frac{2h_t}{LC_p^2}}}.$$

Demostración. Restando (2.18) de (2.17), escogiendo $v = \mathbf{e}_{k+1}^{n+1}$ y después de algunas operaciones algebraicas se deduce:

$$\frac{L}{2} \|\mathbf{e}_{k+1}^{n+1}\|^2 + \frac{L}{2} \|\mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^n\|^2 + h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \\
\leq \frac{L}{2} \|\mathbf{e}_{k+1}^n\|^2 + \langle b(u_{k+1}) - b(u_{k+1}^n), \mathbf{e}_{k+1}^{n+1} - \mathbf{e}_{k+1}^n \rangle \\
- \langle b(u_{k+1}^n) - b(u_{k+1}), u_{k+1}^n - u_{k+1} \rangle.$$

Usando las desigualdades Young y Hölder obtenemos

$$L \|\mathbf{e}_{k+1}^{n+1}\|^{2} + 2h_{t} \|\nabla \mathbf{e}_{k+1}^{n+1}\|^{2} \leq L \|\mathbf{e}_{k+1}^{n}\|^{2} + \frac{1}{L} \|b(u_{k+1}) - b(u_{k+1}^{n})\|^{2} -2 \left\langle b(u_{k+1}^{n}) - b(u_{k+1}), u_{k+1}^{n} - u_{k+1} \right\rangle.$$

Usando el Lema 2.3.1 en el lado derecho se sigue

$$L \|\mathbf{e}_{k+1}^{n+1}\|^2 + 2h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le L \|\mathbf{e}_{k+1}^n\|^2 + \left[\frac{L_b - 2L}{L}\right] \left\langle b(u_{k+1}^n) - b(u_{k+1}), u_{k+1}^n - u_{k+1} \right\rangle.$$

Puesto que $2L > L_b$ el último término de la derecha es negativo por lo tanto

$$L \|\mathbf{e}_{k+1}^{n+1}\|^2 + 2h_t \|\nabla \mathbf{e}_{k+1}^{n+1}\|^2 \le L \|\mathbf{e}_{k+1}^n\|^2.$$

De la desigualdad anterior se deducen las dos desigualdades siguientes:

$$\|\mathbf{e}_{k+1}^{n+1}\| \le \alpha \|\mathbf{e}_{k+1}^{n}\|, \tag{2.19}$$

$$\|\nabla \mathbf{e}_{k+1}^{n+1}\| \le \sqrt{\frac{L}{2h_t}} \|\mathbf{e}_{k+1}^n\|.$$
(2.20)

La desigualdad (2.19) se obtiene aplicando la desigualdad de Poincaré en el segundo término del lado izquierdo. La desigualdad (2.19) demuestra la convergencia del esquema (2.18) en el espacio $L^2(\Omega)$ y (2.20) prueba la convergencia en $H^1(\Omega)$

2.3.2 L-scheme para la forma mixta

El *L-scheme* para la ecuación (2.5) fue estudiado en [4] para una discretización espacial de elementos finitos de Galerkin y un Esquema de Euler totalmente implícito para el tiempo, aquí es necesario imponer algunas restricciones. El mismo problema fue estudiado en [15], pero con un Esquema de Euler semi-implícito en el tiempo.

En (2.5) no usamos ninguna regularización, utilizamos un Esquema de Euler totalmente implícito y elementos finitos de Galerkin en el espacio, obteniendo el problema totalmente discreto siguiente:

 $\begin{cases} \text{Dado } \Psi_k \in V_h, \text{ hallar } \Psi_{k+1} \in V_h \text{ tal que:} \\ \langle \theta(\Psi_{k+1}), v \rangle + h_t \langle K(\theta(\Psi_{k+1})) \left[\nabla \Psi_{k+1} + \vec{e_z} \right], \nabla v \rangle = \langle \theta(\Psi_k) + h_t S_{k+1}, v \rangle \quad (2.21) \\ \text{para cada } v \in V_h \end{cases}$

Asumimos que las funciones en (2.5) satisfacen las hipótesis (M1), (M3), (M5) y (M6). Además en (2.21) se satisfacen las siguientes hipótesis que pueden ser consultadas en [4].

Hipótesis 2.3.1. Suponemos que

- (M7) Existe $L_K > 0$ tal que: $|K(x) - K(y)| \le L_K |x - y|$ para todo $x, y \in \mathbb{R}$.
- (M8) Existe M > 0 tal que $\|\nabla \Psi_k\|_{\infty} \leq M < +\infty$, donde Ψ_k es la solución del problema discreto (2.21).
 - El *L-scheme* para el problema no lineal (2.21) es el siguiente [4]:

$$\begin{cases} \text{Dado } \Psi_k \neq \Psi_{k+1}^n \in V_h, \text{ hallar } \Psi_{k+1}^{n+1} \in V_h \text{ tal que:} \\ L \left\langle \Psi_{k+1}^{n+1} - \Psi_{k+1}^n, v \right\rangle + \left\langle \theta_{k+1}^n, v \right\rangle + h_t \left\langle K(\theta(\Psi_{k+1}^n)) \left[\nabla \Psi_{k+1}^{n+1} + \vec{e_z} \right], \nabla v \right\rangle \\ = \left\langle \theta_k + h_t S_{k+1}, v \right\rangle \text{ para cada } v \in V_h. \end{cases}$$
(2.22)

El elemento clave en el método *L*-scheme, es la adición del término de estabilización $L \langle \Psi_{k+1}^{n+1} - \Psi_{k+1}^{n}, v \rangle$. En lo que se sigue enunciamos el teorema de convergencia del esquema (2.22) el cual puede ser consultado en [4].

Teorema 2.3.3. Asumimos (M1), (M3), (M7), (M8) y la constante L y el paso de tiempo h_t satisfacen la siguiente relación:

$$\frac{2}{L_{\theta}} - \frac{1}{L} - \frac{h_t (M+1)^2 L_K^2}{K_m} \ge 0, \qquad (2.23)$$

entonces el esquema (2.22) converge linealmente con una tasa de convergencia de

$$\sqrt[2]{\frac{1}{1 + \frac{h_t K_m}{L C_p^2}}}.$$
 (2.24)

Demostración. Restando (2.22) de (2.21) se tiene:

$$\begin{split} \left\langle \theta(\Psi_{k+1}^{n}) - \theta(\Psi_{k+1}), v \right\rangle + L \left\langle e_{k+1}^{n+1} - e_{k+1}^{n}, v \right\rangle \\ + h_t \left\langle K(\theta(\Psi_{k+1}^{n})) \nabla \Psi_{k+1}^{n+1} - K(\theta(\Psi_{k+1})) \nabla \Psi_{k+1}, \nabla v \right\rangle \\ + h_t \left\langle \left[K(\theta(\Psi_{k+1}^{n})) - K(\theta(\Psi_{k+1})) \right] e_z, \nabla v \right\rangle = 0. \end{split}$$

Escogiendo $v=e_{k+1}^{n+1}$ y realizando algunas operaciones algebraicas en el tercer y cuarto términos, se deduce que

$$\begin{split} \left\langle \theta(\Psi_{k+1}^{n}) - \theta(\Psi_{k+1}), e_{k+1}^{n+1} \right\rangle + \frac{L}{2} \|e_{k+1}^{n+1}\|^2 + \frac{L}{2} \|e_{k+1}^{n+1} - e_{k+1}^{n}\|^2 - \frac{L}{2} \|e_{k+1}^{n}\|^2 \\ + h_t \left\langle K(\theta(\Psi_{k+1}^{n})) \nabla e_{k+1}^{n+1}, \nabla e_{k+1}^{n+1} \right\rangle \\ + h_t \left\langle \left[K(\theta(\Psi_{k+1}^{n})) - K(\theta(\Psi_{k+1})) \right] \left(\nabla \Psi_{k+1} + e_z \right), \nabla e_{k+1}^{n+1} \right\rangle = 0. \end{split}$$

Usando la hipótesis (M3) obtenemos

$$\begin{aligned} \frac{L}{2} \|e_{k+1}^{n+1}\|^2 + \frac{L}{2} \|e_{k+1}^{n+1} - e_{k+1}^n\|^2 + h_t K_m \|\nabla e_{k+1}^{n+1}\|^2 &\leq \frac{L}{2} \|e_{k+1}^n\|^2 \\ &- \left\langle \theta(\Psi_{k+1}^n) - \theta(\Psi_{k+1}), e_{k+1}^{n+1} - e_{k+1}^n \right\rangle - \left\langle \theta(\Psi_{k+1}^n) - \theta(\Psi_{k+1}), \Psi_{k+1}^n - \Psi_{k+1} \right\rangle \\ &- h_t \left\langle \left[K(\theta(\Psi_{k+1}^n)) - K(\theta(\Psi_{k+1})) \right] (\nabla \Psi_{k+1} + e_z), \nabla e_{k+1}^{n+1} \right\rangle = 0. \end{aligned}$$

Usando el Lema 2.3.1, la hipótesis (M8)y las desigualdades Young y Hölder conseguimos

$$\begin{split} \frac{L}{2} \|e_{k+1}^{n+1}\|^2 + \frac{L}{2} \|e_{k+1}^{n+1} - e_{k+1}^n\|^2 + h_t K_m \|\nabla e_{k+1}^{n+1}\|^2 &\leq \frac{L}{2} \|e_{k+1}^n\|^2 \\ &+ \frac{1}{2L} \|\theta(\Psi_{k+1}^n) - \theta(\Psi_{k+1})\|^2 + \frac{L}{2} \|e_{k+1}^{n+1} - e_{k+1}^n\|^2 \\ &- \frac{1}{L_{\theta}} \|\theta(\Psi_{k+1}^n) - \theta(\Psi_{k+1})\|^2 \\ &+ \frac{h_t (M+1)^2}{2K_m} \|K(\theta(\Psi_{k+1}^n)) - K(\theta(\Psi_{k+1}))\|^2 + \frac{h_t K_m}{2} \|\nabla e_{k+1}^{n+1}\|. \end{split}$$

Simplificando y usando la hipótesis (M7) obtenemos

$$L \|e_{k+1}^{n+1}\|^{2} + h_{t}K_{r} \|\nabla e_{k+1}^{n+1}\|^{2} + \left[\frac{2}{L_{\theta}} - \frac{1}{L} - \frac{h_{t}(M+1)^{2}L_{K}^{2}}{K_{m}}\right] \|\theta(\Psi_{k+1}^{n}) - \theta(\Psi_{k+1})\|^{2} \le L \|e_{k+1}^{n}\|^{2}.$$

De la desigualdad (2.23) se tiene que:

$$L \left\| e_{k+1}^{n+1} \right\|^2 + h_t K_r \left\| \nabla e_{k+1}^{n+1} \right\|^2 \le L \left\| e_{k+1}^n \right\|^2.$$

De aquí se deducen las siguientes desigualdades

$$\left\|\nabla e_{k+1}^{n+1}\right\| \le \sqrt{\frac{L}{h_t K_r}} \left\|e_{k+1}^n\right\|,$$
(2.25)

$$\left\| e_{k+1}^{n+1} \right\| \le \sqrt[2]{\frac{1}{1 + \frac{h_t K_r}{LC_p^2}}} \left\| e_{k+1}^n \right\|.$$
(2.26)

La segunda desigualad se deduce de la desigual
dad de Poincaré, como siempre C_p es la constante de Poincaré.

Observación 2.3.4. La desigualdad (2.26) demuestra que el esquema (2.22), converge linealmente en $L^2(\Omega)$, y la desigualdad (2.25) prueba que también es convergente en el espacio $H^1(\Omega)$. La convergencia de este esquema es global, por lo que es independiente de la solución inicial escogida para iniciar las iteraciones y además no es necesaria una regularización de la función θ . Sin embargo para asegurar la convergencia el paso de tiempo debe cumplir con la siguiente condición que se deriva de la desigualdad (2.23)

$$h_t \leq \left[\frac{2L - L_{\theta}}{L}\right] \left[\frac{K_m}{L_{\theta}(M+2)^2 L_K^2}\right].$$

Es claro que debemos seleccionar $L > \frac{1}{2}L_{\theta}$, sin embargo tomar valores de L muy cercanos a $\frac{1}{2}L_{\theta}$, trae como consecuencia pasos de tiempo muy pequeños. Tomar valores de L muy grandes tampoco es buena opción, ya que $L^{-1}(2L-L_{\theta}) < 2$ y esto no permite que los pasos de tiempo sean arbitrariamente grandes, más aún la tasa de convergencia (2.24) es cercana a 1 lo que deriva en una convergencia lenta. En [4] se recomienda tomar $L = L_{\theta}$ lo que según los experimentos numéricos realizados en este articulo podría ser una buena opción.

De (2.24) se deduce que pasos de tiempo grandes y valores de L pequeños producen tasas de convergencia pequeñas, lo que deriva en una convergencia mas rápida, sin embargo los dos parámetros están ligados por (2.23), donde valores de L pequeños tienen como consecuencia valores de h_t pequeños.

En lo que sigue presentamos el *L-scheme* para forma mixta pero esta vez usando un Esquema de Euler semi-implícito para el tiempo. Asumimos las hipótesis (M1), (M3), (M5), (M6) y (M7) en (2.5). El *L-scheme* para el problema totalmente discreto (2.6) está dado por la siguiente expresión [15].

$$\begin{cases}
\text{Dado } \Psi_k \text{ y } \Psi_{k+1}^n \text{ hallar } \Psi_{k+1}^{n+1} \text{ tal que:} \\
L \left\langle \Psi_{k+1}^{n+1} - \Psi_{k+1}^n, v \right\rangle + \left\langle \theta_{k+1}^n, v \right\rangle + h_t \left\langle K(\Psi_k) \left[\nabla \Psi_{k+1}^{n+1} + \vec{e_z} \right], \nabla v \right\rangle \\
= \left\langle \theta_k + h_t S_{k+1}, v \right\rangle \text{ para cada } v \in V_h.
\end{cases}$$
(2.27)

Teorema 2.3.4. Asumiendo que se cumplen las hipótesis (M1) y (M3) y (M7), y que la constante $L \geq \frac{1}{2}L_{\theta}$, entonces el esquema (2.27) converge linealmente con una tasa de convergencia igual a:

$$\sqrt[2]{\frac{1}{1 + \frac{2h_t K_m}{LC_p^2}}}.$$
 (2.28)

Demostración. Siguiendo las ideas expuestas para la prueba del Teorema 2.3.3 se consigue la siguiente desigualdad:

$$\left\|e_{k+1}^{n+1}\right\|^{2} + \frac{2h_{t}K_{r}}{L}\left\|\nabla e_{k+1}^{n+1}\right\|^{2} \le \left\|e_{k+1}^{n}\right\|^{2},$$

de donde se deduce

$$\left\|\nabla e_{k+1}^{n+1}\right\| \le \sqrt{\frac{L}{2h_t K_r}} \left\|e_{k+1}^n\right\|,$$
(2.29)

$$\left\| e_{k+1}^{n+1} \right\| \le \sqrt[2]{\frac{1}{1 + \frac{2h_t K_r}{LC_p^2}}} \left\| e_{k+1}^n \right\|.$$
(2.30)

De (2.30) se deduce directamente la convergencia en $L^2(\Omega)$, y de (2.29) la convergencia en $H^1(\Omega)$, lo que completa la demostración.

Observación 2.3.5. Es importante destacar que al usar un Esquema de Euler semiimplícito, no es necesaria ninguna restricción en el paso de tiempo. Así, el esquema (2.27) es convergente para cualquier paso de tiempo y cualquier solución inicial escogida. El valor óptimo para el parámetro L es $\frac{1}{2}L_{\theta}$, pues con este valor se obtiene una tasa de convergencia más pequeña en (2.28), así también se puede obtener tasas de convergencia más pequeñas para pasos de tiempo grandes.

2.3.3 L-scheme cuando θ es Hölder continua

En todos los casos que hemos analizado, una hipótesis de suma importancia para probar la convergencia del *L-scheme*, es la Lipschitz continuidad de θ . En [43] se plantea relajar esta hipótesis aceptando únicamente la Hölder continuidad de la función. Estos autores no tratan específicamente la Ecuación de Richards, sino una versión muy simplificada, usan un Esquema de Euler totalmente implícito para el tiempo y el método elementos finitos mixtos para el espacio. Basados en este trabajo [43], presentamos una demostración de la convergencia del esquema (2.27), remplazando la hipótesis de Lipschitz continuidad de θ por Hölder continuidad únicamente, usamos el método de elementos finitos de Galerkin en el espacio.

Vamos a asumir que las funciones involucradas en (2.27) satisfacen las hipótesis (M3)-(M6) y además (M1) es remplazada por

Hipótesis 2.3.2. [43]

(M9) $\theta \in C^1(\mathbb{R}), \ \theta'(z) \ge 0 \ y \ además \ existen \ constantes \ L_{\theta} > 0 \ y \ \gamma \in (0,1) \ tal \ que:$ $|\theta(x) - \theta(y)| \le L_{\theta}|x - y|^{\gamma} \ para \ todo \ x, y \in \mathbb{R}.$

Antes de probar el teorema enunciaremos unos lemas importantes para la demostración.

Lema 2.3.3. Para cualquier C, B, números reales positivos y $\gamma \in (0,1)$ se tiene que la siguiente desigualdad:

$$CB^{1+\gamma} \le \frac{L}{2}B^2 + \frac{1-\gamma}{2}C^{\frac{2}{1-\gamma}}(1+\gamma)^{\frac{1+\gamma}{1-\gamma}} \left(\frac{1}{L}\right)^{\frac{1+\gamma}{1-\gamma}},$$

donde L > 0.

Demostración. Puesto que L > 0 se tiene:

$$CB^{1+\gamma} = \left[B^{1+\gamma} \left(\frac{L}{1+\gamma}\right)^{\frac{1+\gamma}{2}}\right] \left[C \left(\frac{1+\gamma}{L}\right)^{\frac{1+\gamma}{2}}\right]^{1+\gamma}$$

Dado que $\gamma \in (0, 1)$ podemos usar la desigualdad Young con $p = \frac{2}{1+\gamma} \ge 1$ y $q = \frac{2}{1-\gamma}$ de donde se deduce inmediatamente el resultado.

Lema 2.3.4. Para cualesquier A, B > 0 y $\gamma \in (0, 1)$ se verifica la siguiente desigualdad:

$$AB \le L_{\theta}^{-\frac{1}{\gamma}} A^{\frac{1+\gamma}{\gamma}} + \frac{L}{2} B^2 + \frac{1-\gamma}{2} \left[\gamma^{\gamma} L_{\theta} \right]^{\frac{2}{1-\gamma}} (1+\gamma)^{-\frac{1+\gamma}{1-\gamma}} \left(\frac{1}{L} \right)^{\frac{1+\gamma}{1-\gamma}},$$

donde $L_{\theta}, L > 0.$

Demostración. Puesto que $L_{\theta} > 0$ y dado que $\gamma \in (0, 1)$ usamos la desigualdad Young con $p = \frac{1+\gamma}{\gamma} \ge 1$ y $q = 1 + \gamma$ por lo tanto se tiene que:

$$AB \le L_{\theta}^{-\frac{1}{\gamma}} A^{\frac{1+\gamma}{\gamma}} + \left[\frac{\gamma^{\gamma} L_{\theta}}{(1+\gamma)^{1+\gamma}}\right] B^{1+\gamma}.$$

Usando el Lema 2.3.3 en el segundo término de la desigualdad con

$$C = \frac{\gamma^{\gamma} L_{\theta}}{(1+\gamma)^{1+\gamma}},$$

se concluye el resultado.

Lema 2.3.5. Para $v_1, v_2 \in L^2(\Omega)$ y $\gamma \in (0, 1)$ se verifica la siguiente desigualdad:

$$\langle |v_1|, |v_2| \rangle \le L_{\theta}^{-\frac{1}{\gamma}} \|v_1\|_{\frac{1+\gamma}{\gamma}}^{\frac{1+\gamma}{\gamma}} + \frac{L}{2} \|v_2\|^2 + \frac{1-\gamma}{2} [\gamma^{\gamma} L_{\theta}]^{\frac{2}{1-\gamma}} (1+\gamma)^{-\frac{1+\gamma}{1-\gamma}} \left(\frac{1}{L}\right)^{\frac{1+\gamma}{1-\gamma}} \mu(\Omega),$$

donde $L_{\theta}, L > 0.$

1.1.4

Demostración. Usando el Lema 2.3.4 con $A = |v_1|$ y $B = |v_2|$ y luego integrando sobre el conjunto Ω , se tiene el resultado

Lema 2.3.6. Asumiendo (M9) se tiene la siguiente desigualdad:

$$L_{\theta}^{-\frac{1}{\gamma}} \|\theta(v_1) - \theta(v_2)\|_{\frac{1+\gamma}{\gamma}}^{\frac{1+\gamma}{\gamma}} \le \langle \theta(v_1) - \theta(v_2), v_1 - v_2 \rangle, \quad para \ todo \ v_1, v_2 \in L^2(\Omega).$$

Demostración. Usando (M9) y multiplicando los dos lados de la desigualdad por $|\theta(x) - \theta(y)|^{\gamma}$ se tiene:

$$L_{\theta}^{-\frac{1}{\gamma}} |\theta(x) - \theta(y)|^{\frac{1+\gamma}{\gamma}} \le |\theta(x) - \theta(y)| |x - y| \text{ para todo } x, y \in \mathbb{R}.$$

Usando la monotonía de θ , haciendo $x = v_1$, $y = v_2$ e integrando sobre el conjunto Ω se deduce inmediatamente el resultado.

Lema 2.3.7. Si se verifican las hipótesis (M3) y (M9) se tiene la siguiente desigualdad: 2

$$\|e_{k+1}^{n+1}\|^2 + \frac{h_t K_m}{L} \lambda \|\nabla e_{k+1}^{n+1}\|^2 \le \lambda \|e_{k+1}^n\|^2 + R\lambda \left(\frac{1}{L}\right)^{\frac{1}{1-\gamma}},$$

donde

$$R = (1 - \gamma) \left[\gamma^{\gamma} L_{\theta} \right]^{\frac{2}{1 - \gamma}} (1 + \gamma)^{-\frac{1 + \gamma}{1 - \gamma}} \mu(\Omega) \quad y \quad \lambda = \frac{1}{1 + \frac{h_t K_m}{L C_p^2}}.$$

Demostración. Restando (2.27) de (2.6), Escogiendo $v = e_{k+1}^{n+1}$ y realizando algunas operaciones algebraicas se tiene

$$\frac{L}{2} \|e_{k+1}^{n+1}\|^2 + \frac{L}{2} \|e_{k+1}^{n+1} - e_{k+1}^n\|^2 + h_t K_m \|\nabla e_{k+1}^{n+1}\|^2 + \left\langle \theta(\Psi_{k+1}^n) - \theta(\Psi_{k+1}), \Psi_{k+1}^n - \Psi_{k+1} \right\rangle \\
\leq \frac{L}{2} \|e_{k+1}^n\|^2 + \left\langle |\theta(\Psi_{k+1}^n) - \theta(\Psi_{k+1})|, |e_{k+1}^{n+1} - e_{k+1}^n| \right\rangle.$$

Usando el Lema 2.3.6 y 2.3.5 se sigue:

$$\frac{L}{2} \|e_{k+1}^{n+1}\|^{2} + \frac{L}{2} \|e_{k+1}^{n+1} - e_{k+1}^{n}\|^{2} + h_{t}K_{m} \|\nabla e_{k+1}^{n+1}\|^{2} + L_{\theta}^{-\frac{1}{\gamma}} \|\theta(\Psi_{k+1}^{n}) - \theta(\Psi_{k+1})\|_{\frac{1+\gamma}{\gamma}}^{\frac{1+\gamma}{1+\gamma}} \\
\leq \frac{L}{2} \|e_{k+1}^{n}\|^{2} + L_{\theta}^{-\frac{1}{\gamma}} \|\theta(\Psi_{k+1}^{n}) - \theta(\Psi_{k+1})\|_{\frac{1+\gamma}{\gamma}}^{\frac{1+\gamma}{1+\gamma}} + \frac{L}{2} \|e_{k+1}^{n+1} - e_{k+1}^{n}\|^{2} \\
+ \frac{1-\gamma}{2} [\gamma^{\gamma}L_{\theta}]^{\frac{2}{1-\gamma}} (1+\gamma)^{-\frac{1+\gamma}{1-\gamma}} \left(\frac{1}{L}\right)^{\frac{1+\gamma}{1-\gamma}} \mu(\Omega).$$

Simplificando y usando la desigualdad de Poincaré tenemos:

$$(1 + \frac{h_t K_m}{LC_p^2}) \|e_{k+1}^{n+1}\|^2 + \frac{h_t K_m}{L} \|\nabla e_{k+1}^{n+1}\|^2 \le \|e_{k+1}^n\|^2 + (1 - \gamma) [\gamma^{\gamma} L_{\theta}]^{\frac{2}{1 - \gamma}} (1 + \gamma)^{-\frac{1 + \gamma}{1 - \gamma}} \left(\frac{1}{L}\right)^{\frac{2}{1 - \gamma}} \mu(\Omega),$$

de donde finalmente se obtiene el resultado.

Teorema 2.3.5. Asumiendo las hipótesis (M3) y (M9), se tiene que para todo $\varepsilon > 0$, existen L > 0 y $N_0 \in \mathbb{N}$ tal que:

$$||e_{k+1}^n|| \leq \varepsilon$$
 para todo $n \geq N_0$

Demostración. Del Lema 2.3.7 se tiene que

$$||e_{k+1}^{n+1}||^2 \le \lambda ||e_{k+1}^n||^2 + R\lambda \left(\frac{1}{L}\right)^{\frac{2}{1-\gamma}},$$

de donde se deduce inmediatamente que:

$$\|e_{k+1}^{n+1}\|^2 \le \lambda^{n+1} \|e_{k+1}^0\|^2 + \frac{\lambda}{1-\lambda} R\left(\frac{1}{L}\right)^{\frac{2}{1-\gamma}}$$

Puesto que $0 < \lambda < 1$, dado $\varepsilon > 0$ existe $N_0 \in \mathbb{N}$ tal que

$$\lambda^{n+1} \|e_{k+1}^0\|^2 \le \frac{\sqrt{\varepsilon}}{2} \text{ para todo } n \ge N_0, \qquad (2.31)$$

por lo tanto se puede seleccionar L > 0 suficiente grande tal que:

$$\frac{\lambda}{1-\lambda} R\left(\frac{1}{L}\right)^{\frac{2}{1-\gamma}} \le \frac{\sqrt{\varepsilon}}{2}.$$
(2.32)

Para obtener (2.32) es suficiente escoger

$$L > \left[\frac{2\lambda R}{\sqrt{\epsilon}(1-\lambda)}\right]^{\frac{1}{2}(1-\gamma)}.$$

De las desigualdades (2.31) y (2.32) se concluye inmediatamente el resultado.

Observación 2.3.6. Si θ es Lipschitz continua, el problema se reduce al estudiado en el Teorema 2.3.4. En este caso, la iteración es una contracción, por lo que la convergencia es incondicional y converge para cualquier $L \geq \frac{1}{2}L_{\theta}$.

Observación 2.3.7. Observe que la convergencia se puede lograr sin importar el paso de tiempo h_t . De hecho, cuanto mayor sea el paso de tiempo, mejor será la convergencia del esquema iterativo. La tasa de convergencia empeora para valores de L grandes, ya que λ se acerca a 1. Desde el punto de vista teórico, esto da como resultado un mayor número de iteraciones para obtener la precisión deseada. Sin embargo, esta es una interpretación bastante pesimista, ya que en los experimentos numéricos presentados en [43] el número de iteraciones es razonable.

2.4 Esquemas de linealización mixtos

Los esquemas de *Picard* y *L-scheme* son robustos, pero sólo linealmente convergentes, mientras que el método de *Newton* es cuadráticamente convergente, aunque solo locamente convergente. Se busca combinar la robustez de los métodos de *Picard* y *L-scheme* con la velocidad del método de *Newton*, es por esto que se han propuesto dos métodos: el *Picard/Newton* y el *L-esquema/Newton*. Estos métodos consisten en calcular algunas iteraciones con el esquema robusto antes de cambiar al método de *Newton*. Los nuevos métodos mixtos tienen un mejor rendimiento tanto en robustez como en velocidad de cómputo, sin embargo su convergencia global no ha sido demostrada teóricamente.

El esquema de *Picard/Newton* fue propuesto por [14] y es una combinación del método de *Picard* y el método de *Newton*. Su convergencia no está asegurada puesto que comenzamos las iteraciones con el esquema de *Picard* y por el Teorema 2.2.1 este converge sólo localmente, es decir, si el método de *Picard Modificado* falla también falla el esquema de *Picard/Newton*.

El esquema *L-esquema/Newton* fue propuesto en [4] y consiste en hacer un número fijo de iteraciones con el *L-scheme* y luego cambiar al método de *Newton*. En [4] se recomienda hacer entre 4 o 5 iteraciones del *L-scheme* antes de cambiar al método de *Newton*. La convergencia de este esquema tampoco está asegurada, puesto que no se puede asegurar que luego de un número fijo de iteraciones del *L-scheme*, la solución obtenida esté suficientemente cerca de la solución exacta para garantizar la convergencia del método de *Newton*.

Capítulo 3

UNA FAMILIA DE NUEVOS ESQUEMAS DE LINEALIZACIÓN

En este capítulo buscamos una generalización del *L*-scheme para la Ecuación de Richards. Construimos una familia de esquemas de linealización de primer orden robustos, eficientes y globalmente convergentes. Usamos una sucesión de funciones reales L^n , $n \in \mathbb{N}$, para construir un marco teórico general. Para la discretización espacial se usa el método de elementos finitos de Galerkin linealeales. La discretización del tiempo se basa en el método de Euler semi-implícito. Obtenemos cinco esquemas que mejoran la convergencia del *L*-scheme: Derivative Globally convergent Linear Scheme (DGLS), Globally convergent Linear Scheme (GLS), Modified Newton's Scheme (MNS), Modified Derivative Globally convergent Linear Scheme (MDGLS) y Modified Globally convergent Linear Scheme (MGLS). Demostramos la convergencia de las iteraciónes en la norma $H^1(\Omega)$ y verificamos los resultados teóricos mediante tres ejemplos numéricos. Los nuevos métodos son comparados con el *L*-scheme y el método de Newton, estudiamos el orden de convergencia, el número de iteraciones, y tiempo de de CPU de los métodos.

El MNS es un método más que robusto que el esquema de Newton aunque los dos tienen el mismo orden de convergencia. Los experimentos numéricos muestran que el MNS converge cuando el esquema de Newton no lo hace y usa un tiempo de cómputo similar, lo que lo convierte en una buena alternativa. Los cuatro restantes son esquemas globalmente convergentes, con orden de convergencia lineal. Los métodos DGLSy GLS muestran prácticamente el mismo rendimiento, que en general es mejor que el del L-scheme, dado que el esquema DGLS usa la derivada de θ y el GlS no lo hace, este podría tener un extraordinario desempeño como en el caso del Ejemplo 2 con la segunda condición inicial. Para mejorar el rendimiento de estos esquemas se desarrollo dos métodos adicionales, el MDGLS y el MGLS aunque presentan un rendimiento similar, dado que el MDGLS usa la derivada de la función θ su desempeño es escasamente mejor. Los resultados muestran que estos dos nuevos esquemas obtiene las mejores resultados, no solo por su convergencia global, sino que además exhiben mejores tasas de convergencia (hasta cuatro veces mejor que la del *L-scheme*), un menor número de iteraciones y tiempos de CPU más bajos. Es importante destacar que se pueden desarrollar otros métodos de linealización diferentes a los expuestos en este trabajo utilizando este marco teórico.

3.1 Discretización del tiempo

La discretización en el tiempo se logra mediante un esquema Euler explícito-implícito en la Ecuación de Richards en su forma mixta (2.5). Restringimos las formulaciones y el análisis a condiciones de frontera de Dirichlet homogéneas, solo por simplicidad. La extensión a condiciones de frontera más generales es sencilla (los ejemplos numéricos presentados en esta sección involucran otras condiciones de frontera). Denotemos el paso de tiempo $h_t = T/N$ para algún $N \in \mathbb{N}$ fijo, la solución aproximada $\psi_k := \psi(t_k)$, $t_k = kh_t$, y $\theta_k = \theta(\Psi_k)$. Obtenemos el siguiente problema elíptico no lineal con las condiciones de contorno de Dirichlet: para k = 1, 2, ..., N

$$\begin{cases} \text{Dado } \Psi_k \in H_0^1(\Omega), \text{ hallar } \Psi_{k+1} \in H_0^1(\Omega) \text{ tal que:} \\ \langle \theta_{k+1}, \rho \rangle + h_t \langle K(\Psi_k)(\nabla \Psi_{k+1} + \vec{e_z}), \nabla \rho \rangle = \langle \theta_k + h_t S_{k+1}, \rho \rangle, \\ \text{para todo } \rho \in H_0^1(\Omega). \end{cases}$$
(3.1)

En el primer paso de tiempo Ψ_0 es la condición inicial. Se asumen las hipótesis (M1), (M3), (M4), (M5) y (M6) para el problema (3.1). Este esquema de tiempo fue usado por Slodicka [15] y también en [32] con el *L*-scheme en la cuarta forma de la Ecuación de Richards y elementos finitos mixtos.

El método de Newton para (3.1) fue estudiado en el capítulo anterior (ver 2.1.2). Dado que usamos un Esquema de Euler semi-implícito para discretizar el tiempo, el esquema de Newton de (3.1) no involucra la derivada de la función de conductividad hidráulica K, ya que se evalua en el tiempo anterior y por lo tanto no constituye un término no lineal en la ecuación. Este procedimiento es diferente al propuesto por Celia et al [3], donde se usa un Esquema de Euler totalmente implícito obteniendo dos términos no lineales. La no linealidad que involucra la derivada en el tiempo se discretiza cuadráticamente mientra que la no linealidad de K se aproxima linealmente. A este método lo denominamos esquema de Picard Modificado y fue estudiado en el capítulo anterior (ver 2.2).

3.2 Linealización y estimación del Error

En esta sección proponemos el siguiente esquema general para la linealización de (3.1): Para k=1,2,...,N

$$\begin{cases} \operatorname{dado} \Psi_k, \Psi_{k+1}^n \in H_0^1(\Omega), \operatorname{hallar} \Psi_{k+1}^{n+1} \in H_0^1(\Omega) \text{ tal que:} \\ \left\langle \widehat{L}_{k+1}^n \left(\Psi_{k+1}^{n+1} - \Psi_{k+1}^n \right), \rho \right\rangle + h_t \left\langle K(\Psi_k) \left(\nabla \Psi_{k+1}^{n+1} + \vec{e_z} \right), \nabla \rho \right\rangle \\ = \left\langle \theta_k - \theta_{k+1}^n + h_t S_{k+1}, \rho \right\rangle \\ \text{para todo } \rho \in H_0^1(\Omega), \text{ y para } n = 0, 1, 2, \dots, \end{cases}$$
(3.2)

donde $\widehat{L}_{k+1}^n : \Omega \to \mathbb{R}^+$ es positiva. Observemos que el esquema (3.2) tiene una solución única $\Psi_{k+1}^{n+1} \in H_0^1(\Omega)$ para $k \neq n$ fijos. Esto se sigue inmediatamente de la teoría de las ecuaciones elípticas lineales [59, teorema 3 pagina 301] y los supuestos (M1) y (M3)-(M6). Tenga en cuenta que si existe γ tal que $0 < \gamma \leq \widehat{L}_{k+1}^n$ casi en todo punto de Ω , podemos definir un producto interno ponderado en $L^2(\Omega)$ y su norma inducida de la siguiente manera:

$$\langle f,g \rangle_n = \int_{\Omega} \widehat{L}_{k+1}^n fg \, dx, \qquad \|f\|_n = \left(\int_{\Omega} \widehat{L}_{k+1}^n |f|^2\right)^{\frac{1}{2}}.$$

En lo que sigue introducimos algunas hipótesis sobre \widehat{L}_{k+1}^n .

Hipótesis 3.2.1. Sean $\epsilon, \beta > 0$. Definimos para todo $x \in \Omega$ fijo y n = 1, 2, ... $I_{k+1}^n(x) := \left\{ z = \omega \Psi_{k+1}^n(x) + (1-\omega) \Psi_{k+1}(x), \ \omega \in [0,1] \right\} \subset \mathbb{R}$

(H10) $\widehat{L}_{k+1}^n(x) = C \ge \frac{L_{\theta}}{2}$, donde C es una constante positiva.

- (H11) $0 < \frac{1}{2}L_{\theta} \le \widehat{L}_{k+1}^n(x) \le \beta$ para todo $x \in \Omega$ $y \ n = 1, 2, ...$
- (H12) $0 < \epsilon \leq \frac{1}{2} \max \left\{ \frac{\theta'(z)}{z \in I_{k+1}^n(x)} \right\} \leq \widehat{L}_{k+1}^n(x) \leq \beta \text{ para todo } x \in \Omega \ y \ n = 1, 2, \dots$
- (H13) $\widehat{L}_{k+1}^n(x) \leq \widehat{L}_{k+1}^{n-1}(x)$ para todo $x \in \Omega$ y n = 1, 2, ...

Observación 3.2.1. Si la función θ' no es estrictamente mayor a ε previo a la discretización, podemos usar un paso de regularización en (2.5). Específicamente, dado $\epsilon > 0$, definimos la función no lineal θ_{ϵ} como una aproximación de la no linealidad original θ satisfaciendo $\theta'_{\epsilon} > \epsilon$. Para mayores detalles podemos consultar [32,65]. Con la finalidad de simplificar, se puede considerar la perturbación global $\theta_{\epsilon}(z) = \theta(z) + \epsilon z$.

Ahora enunciamos el principal resultado de este capítulo que lo usaremos posteriormente para construir nuevos esquemas. **Teorema 3.2.1.** Suponiendo que se cumple (M3) entonces se verifica:

a) Para k = 1, 2, ..., N, y n = 0, 1, ...

$$\|e_{k+1}^{n+1}\|_{n}^{2} + h_{t}K_{r}\|\nabla e_{k+1}^{n+1}\|^{2} \leq \left\langle \int_{I_{k+1}^{n}} \left|\theta'(t) - \widehat{L}_{k+1}^{n}\right| dt, \left|e_{k+1}^{n+1}\right| \right\rangle$$
(3.3)

- b) Asumiendo (M1) y uno de los siguientes tres casos: (M10) o bien (M11) y (M13) o bien (M12) y (M13), entonces, la sucesión de soluciones $\{\Psi_{k+1}^n\}_{n>0}$ de (3.2) converge a Ψ_{k+1} solución de (3.1) en $H_0^1(\Omega)$ para $k \in \{1, ..., N\}$.
- c) Asumiendo (M1) para todo n = 1, 2, ..., y para todo k = 1, 2, ..., N fijo, tenemos la siguiente estimación del error

$$\|e_{k+1}^{n}\|_{H^{1}(\Omega)} \leq \max\left\{C_{p}, 1\right\} \frac{\beta C_{p}}{h_{t}K_{r}} \sqrt{2 + \frac{L_{\theta}h_{t}K_{r}}{2\beta^{2}C_{p}^{2}}} \|\Psi_{k+1}^{n} - \Psi_{k+1}^{n-1}\|.$$
(3.4)

Demostración. : a) Restando (3.1) de (3.2), y eligiendo $\rho = e_{k+1}^{n+1}$, obtenemos

$$\left\langle \widehat{L}_{k+1}^{n} \left(\Psi_{k+1}^{n+1} - \Psi_{k+1}^{n} \right), e_{k+1}^{n+1} \right\rangle + h_t \left\langle K(\Psi_k) \nabla e_{k+1}^{n+1}, \nabla e_{k+1}^{n+1} \right\rangle = \left\langle \theta_{k+1} - \theta_{k+1}^{n}, e_{k+1}^{n+1} \right\rangle.$$

Usando (M3) y haciendo algunas operaciones algebraicas uno consigue

$$\|e_{k+1}^{n+1}\|_n^2 + h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 \le \left\langle \left(\theta_{k+1} - \theta_{k+1}^n\right) - \widehat{L}_{k+1}^n \left(\Psi_{k+1} - \Psi_{k+1}^n\right), e_{k+1}^{n+1} \right\rangle.$$

Usando la definición de I_{k+1}^n , probamos (3.3).

b) Primer caso. De (M1) y (M10) se sigue $|\theta'(t) - C| \leq C$. Usando este resultado en (3.3), obtenemos la siguiente desigualdad

$$C \|e_{k+1}^{n+1}\|^2 + h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 \le C \left\langle \left|e_{k+1}^n\right|, \left|e_{k+1}^{n+1}\right| \right\rangle.$$

Usando las desigualdades de Cauchy-Schwarz y Young, se tiene

$$\|e_{k+1}^{n+1}\|^2 + \frac{2h_t K_r}{C} \|\nabla e_{k+1}^{n+1}\|^2 \le \|e_{k+1}^n\|^2.$$
(3.5)

De la desigualdad de Poincaré con constante C_p , se tiene que

$$\left\|e_{k+1}^{n+1}\right\| \le \left(1 + \frac{2h_t K_r}{C_p^2 C}\right)^{-\frac{1}{2}} \left\|e_{k+1}^n\right\| \quad \mathbf{y} \qquad \left\|\nabla e_{k+1}^{n+1}\right\| \le \sqrt[2]{\frac{C}{2h_t K_r}} \left\|e_{k+1}^n\right\|.$$
(3.6)

De (3.6), se sigue inmediatamente la convergencia del esquema en $H^1(\Omega)$.

Segundo caso: Seguimos las mismas ideas de la prueba anterior. Usando (M11), obtenemos

$$\|e_{k+1}^{n+1}\|_n^2 + h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 \le \left\langle \widehat{L}_{k+1}^n e_{k+1}^n, e_{k+1}^{n+1} \right\rangle.$$
(3.7)

De las desigualdades de Cauchy-Schwarz y Young, se tiene la siguiente desigualdad

$$\|e_{k+1}^{n+1}\|_n^2 + 2h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 \le \|e_{k+1}^n\|_n^2.$$
(3.8)

Usando (M13) y la desigualdad de Poincaré, obtenemos

$$\begin{cases} \left\| e_{k+1}^{n+1} \right\| \leq \frac{2\beta}{L_{\theta}} \left[\sqrt{\frac{1}{1 + \frac{2h_{t}K_{r}}{C_{p}^{2}\beta}}} \right]^{n+1} \left\| e_{k+1}^{0} \right\|, \\ \left\| \nabla e_{k+1}^{n+1} \right\| \leq \sqrt{\frac{\beta}{2h_{t}K_{r}}} \left\| e_{k+1}^{n} \right\|. \end{cases}$$
(3.9)

De (3.9), se obtiene directamente la convergencia del esquema en $H^1(\Omega)$.

Tercer caso: De (M12) se deduce fácilmente que para cada $x \in \Omega$ fijo, se verifica $|\theta'(t) - \hat{L}_{k+1}^n(x)| \leq \hat{L}_{k+1}^n(x)$ para todo $t \in I_{k+1}^n(x)$. Usando este resultado en (3.3) se obtiene (3.7) y luego (3.8). Luego usando (M13) se tiene la siguiente desigualdad

$$\left\| e_{k+1}^{n+1} \right\| \le \frac{\beta}{\epsilon} \left(1 + \frac{2h_t K_r}{C_p^2 \beta} \right)^{-\frac{n+1}{2}} \left\| e_{k+1}^0 \right\|, \tag{3.10}$$

y de (3.8), se sigue que

$$2h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 \le \|e_{k+1}^n\|_n^2.$$

De la desigual dad de Poincaré, la expresión $(3.10), \, {\rm y}$ después de algun as simplificaciones, obtenemos

$$\|e_{k+1}^{n+1}\|_{H^1(\Omega)} \le \left(1 + \frac{2h_t K_r}{C_p^2 \beta}\right)^{-\frac{n}{2}} \max\left\{C_p, 1\right\} \sqrt{\frac{L_\theta}{h_t K_r}} \|e_{k+1}^0\|.$$
(3.11)

De lo anterior se tiene directamente la convergencia.

c) Restando (3.1) de (3.2), escogiendo $\rho = e_{k+1}^{n+1}$ y usando (M3) obtenemos

$$h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 + \left\langle \theta_{k+1}^n - \theta_{k+1}, \Psi_{k+1}^n - \Psi_{k+1} \right\rangle \leq \left\langle \Psi_{k+1}^n - \Psi_{k+1}^{n+1}, e_{k+1}^{n+1} \right\rangle_n + \left\langle \theta_{k+1} - \theta_{k+1}^n, \Psi_{k+1}^{n+1} - \Psi_{k+1}^n \right\rangle.$$
(3.12)

Por otro lado de (M1) se puede deducir la siguiente desigualdad

$$\frac{1}{L_{\theta}} \|\theta_{k+1}^{n} - \theta_{k+1}\|^{2} \le \left\langle \theta_{k+1}^{n} - \theta_{k+1}, \Psi_{k+1}^{n} - \Psi_{k+1} \right\rangle.$$
(3.13)

Usando (3.13) en (3.12) obtenemos

$$\frac{1}{L_{\theta}} \|\theta_{k+1}^{n} - \theta_{k+1}\|^{2} + h_{t}K_{r} \|\nabla e_{k+1}^{n+1}\|^{2} \leq \beta \|\Psi_{k+1}^{n} - \Psi_{k+1}^{n+1}\| \|e_{k+1}^{n+1}\| + \|\theta_{k+1} - \theta_{k+1}^{n}\| \|\Psi_{k+1}^{n+1} - \Psi_{k+1}^{n}\| + \|\theta_{k+1} - \theta_{k+1}^{n}\| + \|\theta_{k+1} - \theta_{k+1}^{n}\| \|\Psi_{k+1}^{n+1} - \Psi_{k+1}^{n}\| + \|\theta_{k+1} - \theta_{k+1}^{n}\| + \|\theta_{k+1}^{n}\| + \|\theta_{k+1} - \theta_{k+1}^{n}\| + \|\theta_{k+1} - \theta_{k+1}^{n}\| + \|\theta_{k$$

Usando las desigualdades de Poincaré y Young se tiene

$$\frac{h_t K_r}{4C_p^2} \|e_{k+1}^{n+1}\|^2 + \frac{h_t K_r}{2} \|\nabla e_{k+1}^{n+1}\|^2 \le \left[\frac{\beta^2 C_p^2}{h_t K_r} + \frac{L_\theta}{4}\right] \|\Psi_{k+1}^{n+1} - \Psi_{k+1}^n\|^2.$$

De la desigualdad anterior finalmente obtenemos el resultado.

Observación 3.2.2. Si $\Omega \subset \mathbb{R}$, usando el Lema 4.3.4 de [52], obtenemos la desigualdad $\|e_{k+1}^{n+1}\|_{L^{\infty}(\Omega)} \leq \|e_{k+1}^{n+1}\|_{H^{1}_{0}(\Omega)}$. Esta desigualdad se puede utilizar en las estimaciones (3.4) y (3.11), lo que permite obtener estimaciones en $L^{\infty}(\Omega)$ al menos para una dimensión.

3.3 Esquema de linealización

Formulamos varios esquemas numéricos basados en el esquema general propuesto y discutido en la sección 3.2. Estos esquemas nuevos son robustos y sus convergencias son de primer o segundo orden. Estos esquemas involucran funciones de aproximación \hat{L}_{k+1}^n más generales que la función constante usada para definir el *L-scheme*, de hecho el *L-scheme* es un caso particular de estos métodos. Téngase en cuenta que es posible desarrollar otros esquemas diferentes a los propuestos en esta tesis usando la teoría expuesta en el capítulo anterior basados, por ejemplo, en otras aproximaciones de la derivada de θ .

3.3.1 *L*-scheme y *L*2-scheme

Si $\widehat{L}_{k+1}^n = \theta'(\Psi_{k+1}^n)$, obtenemos el esquema de *Newton* que fue abordado en la sección 2.1.2. Si $\widehat{L}_{k+1}^n(x) = C \ge \frac{L_{\theta}}{2}$, con *C* constante; entonces obtenemos el *L*-scheme que puede ser consultado en [4, 15, 32] y con mayor detalle en la sección 2.3 de esta tesis Su convergencia fue probada en el Teorema 3.2.1 parte b. Si $C = L_{\theta}$, obtenemos *L*-scheme propuesto en [15] mientras que si consideramos $C = L_{\theta}/2$ encontramos el esquema sugerido en [4] y lo notamos como *L*2-scheme.

3.3.2 Esquemas DGLS y GLS

Usando el Teorema 3.2.1 y las hipótesis (M11) y (M13), definiremos los dos primeros esquemas de linealización, estos son de primer orden y globalmente convergentes. Sabemos que el método de Newton tiene un mejor rendimiento que el *L*-scheme, aunque solo converge localmente. Es por esto que parece natural elegir \hat{L}_{k+1}^n como



Figura 3.1: Funciones L_1 y L_2 para los esquemas DGLS y GLS.

una aproximación de la función $\theta'(\Psi_{k+1}^n)$, de tal manera que satisfaga las hipótesis requeridas en el Teorema 3.2.1, y así conseguir la convergencia global del método. Con estas hipótesis, obtenemos un mejor rendimiento que el *L*-scheme.

Definimos dos nuevos esquemas numéricos con orden de convergencia lineal, el primer esquema usa la derivada de la función θ , a este lo llamamos **Derivative Globally convergent Linear Scheme (DGLS)**. Este esquema toma los valores de $\theta'(\Psi_{k+1}^n)$ si $\theta'(\Psi_{k+1}^n) \geq \frac{1}{2}L_{\theta}$ y un valor constante igual a $\frac{1}{2}L_{\theta}$ en caso contrario, lo que garantiza la convergencia del esquema. El segundo esquema lo denominamos **Globally convergent Linear Scheme (GLS)** y no usa θ' , lo que de alguna manera significa una ventaja. Este esquema está inspirado en el ejemplo numérico 2 presentado en esta tesis, donde se consideraron dos condiciones iniciales diferentes. Al analizar la convergencia de los métodos *L-scheme* y *L2-scheme*, se puede observar que, usando la primera condición inicial con $L = \frac{1}{2}L_{\theta}$, se obtuvo una convergencia más rápida, mientras que con la segunda condición inicial se obtuvo una convergencia más rápida para $L = L_{\theta}$. Estos resultados sugieren un nuevo esquema que involucre estos dos valores para aproximar la función θ' .

Definimos $L_1(z)$ y $L_2(z)$ como se muestra en (3.14) y (3.15) respectivamente, ver adicionalmente también la Figura 3.1.

$$L_1(z) = \begin{cases} \theta'(z), & \text{si } \theta'(z) \ge \frac{L_{\theta}}{2}, \\ \frac{L_{\theta}}{2}, & \text{si } \theta'(z) < \frac{L_{\theta}}{2}, \end{cases}$$
(3.14)

$$L_2(z) = \begin{cases} L_{\theta}, & \text{si } |z - \hat{z}| < \hat{h}, \\ \frac{L_{\theta}}{2}, & \text{si } |z - \hat{z}| \ge \hat{h}. \end{cases}$$
(3.15)

Denotamos \hat{z} el único punto máximo de la función θ' , para algunos detalles podemos referirnos la sección 1.1.5 de esta tesis. El parámetro \hat{h} se selecciona adecuadamente, de tal manera que $L_2(z)$ esté cerca de θ' como se muestra en Figura 3.1. Se puede seleccionar \hat{h} de muchas maneras diferentes y para los experimentos numéricos presentados en esta tesis, seleccionamos $\hat{h} = \frac{1}{2}|x_1 - x_2|$, donde x_1, x_2 se encuentran resolviendo la ecuación $\theta'(x) = \frac{3}{4}L_{\theta}$. Sin embargo, existen otras maneras de seleccionar \hat{h} . Esta aproximación puede verse en la Figura 3.1.

Para determinar los nuevos esquemas numéricos usamos un parámetro $r \in \mathbb{N}$, y a continuación definimos la sucesión $\hat{L}_{k+1}^n = L_m(\psi_{k+1}^n)$ para todo n = 0, 1, 2, ..., r y $\hat{L}_{k+1}^n = L_m(\psi_{k+1}^r)$ para n = r+1, r+2, ..., con m = 1 para el esquema *DGLS* y m = 2para el esquema *GLS*. Manteniendo \hat{L}_{k+1}^n constante para n > r, nos aseguramos que se satisface la hipótesis (*M13*) y conseguimos que se cumplan los supuestos necesarios para garantizar la convergencia de los esquemas. Es importante destacar, que al usar hipótesis (*M11*), no es necesario usar una regularización de la función θ , aun cuando la derivada de esta función se anule en parte de su dominio.

3.3.3 Esquema *MNS*

Puesto que los esquemas DGLS y GLS no proveen una buena aproximación de θ' para valores mas pequeños que $\frac{1}{2}L_{\theta}$, formulamos un nuevo esquema que corrige esta desventaja llamado **Modified Newton's Scheme (MNS)**. Este nuevo método usa la estimación del error en $L^{\infty}(\Omega)$ dada en (3.4) para desarrollar un esquema robusto y en muchos casos de segundo orden. En lo que se sigue suponemos que existe $\varepsilon > 0$ tal que $\theta' > \varepsilon$. En el caso que esta propiedad no se cumpla, se debe usar una regularización de la función como se indicó en la Observación 3.2.1. Con la finalidad de aproximar \hat{L}_{k+1}^n a $\theta'(\Psi_{k+1}^n)$ en cada iteración, definimos:

$$d_{k+1}^n := C_e \|\Psi_{k+1}^{n+1} - \Psi_{k+1}^n\| \quad \text{y} \quad E_{k+1}^n(x) := \frac{1}{2} \max\left\{\theta'_\epsilon(z) : |z - \Psi_{k+1}^n(x)| \le d_{k+1}^n\right\}.$$

Notamos que θ' tiene un único punto crítico $\hat{z} \leq 0$ y además θ' es monótona creciente en $(-\infty, \hat{z})$ y monótona decreciente en $(\hat{z}, +\infty)$. De la parte anterior se sigue que

 E_{k+1}^n se puede escribir de la siguiente manera:

$$E_{k+1}^{n}(x) = \begin{cases} \frac{1}{2}\theta'(\Psi_{k+1}^{n}(x) + d_{k+1}^{n}), & \text{si } \hat{z} - \Psi_{k+1}^{n}(x) > d_{k+1}^{n}, \\ \frac{1}{2}L_{\theta}, & \text{si } |\hat{z} - \Psi_{k+1}^{n}(x)| \le d_{k+1}^{n}, \\ \frac{1}{2}\theta'(\Psi_{k+1}^{n}(x) - d_{k+1}^{n}), & \text{si } \hat{z} - \Psi_{k+1}^{n}(x) < -d_{k+1}^{n}. \end{cases}$$

Definimos el esquema MNS eligiendo:

$$\widehat{L}_{k+1}^n(x) = \max\left\{E_{k+1}^n(x), \theta_{\epsilon}'(\Psi_{k+1}^n(x))\right\}.$$

Está claro que \widehat{L}_{k+1}^n satisface (M12), por lo tanto podemos obtener la siguiente estimación $||e_{k+1}^{n+1}||_n \leq \lambda ||e_{k+1}^n||_n$, con $0 < \lambda < 1$ constante. Sin embargo no necesariamente \widehat{L}_{k+1}^n satisface (M13) por lo que no podemos asegurar la convergencia del esquema. Después de un numero finito de iteraciones del método MNS, recuperamos el esquema Newton, por lo tanto el esquema tiene el mismo orden de convergencia que el método de Newton.

Definimos la función $S^n_{k+1}:\mathbb{R}\to\mathbb{R}$ siguiente:

$$S_{k+1}^{n}(y) = \begin{cases} \frac{1}{2}\theta'(y+d_{k+1}^{n}), & \text{si } \hat{z}-y > d_{k+1}^{n}, \\\\ \frac{1}{2}L_{\theta}, & \text{si } |\hat{z}-y| \le d_{k+1}^{n}, \\\\ \frac{1}{2}\theta'(y-d_{k+1}^{n}), & \text{si } \hat{z}-y < -d_{k+1}^{n}. \end{cases}$$

Además definimos $L_{k+1}^n(y) = \max \{S_{k+1}^n(y), \theta'(y)\}$. Resulta evidente que $E_{k+1}^n = S_{k+1}^n(\Psi_{k+1}^n)$ y $\widehat{L}_{k+1}^n = L_{k+1}^n(\Psi_{k+1}^n)$. En la Figura 3.4, mostramos un ejemplo de la evolución de L_{k+1}^n en cada iteración. Nótese como en la cuarta iteración, prácticamente L_{k+1}^n ha alcanzado a la función θ' , por lo tanto, a partir de la iteración 4 el esquema MNS coincide con el método de Newton.

3.3.4 Esquemas *MDGLS* **y** *MGLS*

El esquema MNS no satisface el supuesto (M13), por lo que no se puede probar la convergencia global del método, sin embargo los experimentos numéricos presentados, muestran que es un esquema mucho más robusto que el método Newton. Definimos dos nuevos esquemas que corrigen esta desventaja. Recordando los esquemas DGLS y GLS, tratamos de mejor la aproximación de θ' y los llamamos **Modified DGLS**

(MDGLS) y Modified GLS (MGLS). Para formular los esquemas, usamos la estimación del error en $L^{\infty}(\Omega)$ dada en el Teorema 3.11, que evidentemente es monótona decreciente. Notamos que $||e_{k+1}^0||$ se puede calcular usando la estimación demostrada en la Proposición 3.5 de [31]. La idea clave de estos dos esquemas es la construcción de los intervalos $A_{k+1}^n(x) \supset I_{k+1}^n(x)$, lo que permite garantizar la hipótesis (M12). Estos intervalos están encajados, lo que además garantiza la hipótesis (M13). Adicionalmente, usamos las propiedades de θ' descritos en los capítulos anteriores para los modelos de Van-Genuchten y Haverkamp con la finalidad de escribir los esquemas de una manera más sencilla.

Consideramos la sucesión $\{d_{k+1}^n\}_{n=1,2,\ldots} \subset \mathbb{R}$ tal que $\|e_{k+1}^n\|_{L^{\infty}(\Omega)} \leq d_{k+1}^n$, por ejemplo usando el Teorema 3.11, podemos definir $d_{k+1}^n = \left(1 + \frac{2h_t K_r}{C_p^2 \beta}\right)^{-\frac{n-1}{2}} \max\{C_p, 1\} \sqrt{\frac{L_{\theta}}{h_t K_r}}$, para $n = 1, 2, \ldots$ A continuación definimos los intervalos encajados: para todo $n = 1, 2, \ldots$

$$A_{k+1}^n(x) := \left[a_{k+1}^n(x), b_{k+1}^n(x)\right] = \bigcap_{i=1}^n \left[\Psi_{k+1}^i(x) - 2d_{k+1}^i, \Psi_{k+1}^i(x) + 2d_{k+1}^i\right].$$

Es claro que las siguientes propiedades son válidas:

$$A_{k+1}^n(x) \subset A_{k+1}^{n-1}(x), \tag{3.16}$$

si
$$d_{k+1}^n \le d_{k+1}^{n-1} \le \dots \le d_{k+1}^1$$
, entonces $I_{k+1}^n(x) \subset A_{k+1}^n(x)$. (3.17)

Para construir \widehat{L}_{k+1}^n , tomamos $\tau > 0.5$ fijo y definimos $\widehat{L}_{k+1}^0(x) := \tau L_{\theta}$ para todo $x \in \Omega$. De (3.11), se sigue que $\|e_{k+1}^1\|_{L^{\infty}(\Omega)} \leq d_{k+1}^1 := \max\{C_p, 1\} \sqrt{\frac{L_{\theta}}{h_t K_r}} \|e_{k+1}^0\|$ y a continuación, definimos

$$\widehat{L}_{k+1}^{1}(x) := \tau \max\left\{\theta_{\epsilon}'(z) : z \in A_{k+1}^{1}(x)\right\} \le \widehat{L}_{k+1}^{0}(x).$$

Usando de nuevo (3.11), obtenemos que $\|e_{k+1}^2\|_{L^{\infty}(\Omega)} \leq d_{k+1}^2 := \left(1 + \frac{2h_t K_r}{C_p^2 \beta}\right)^{-\frac{1}{2}} d_{k+1}^1$. Por inducción y utilizando d_{k+1}^n podemos definir

$$\widehat{L}_{k+1}^{n}(x) := \tau \max \left\{ \theta'(z) : z \in A_{k+1}^{n}(x) \right\},\$$

que satisface la propiedad de monotonicidad $\widehat{L}_{k+1}^n \leq \widehat{L}_{k+1}^{n-1}$. De (3.11), obtenemos la estimación del error $\|e_{k+1}^{n+1}\|_{L^{\infty}(\Omega)} \leq d_{k+1}^{n+1} := \left(1 + \frac{2h_t K_r}{C_p^2 \beta}\right)^{-\frac{1}{2}} d_{k+1}^{n-1}$, que converge a cero. Observe que la convergencia en $H_0^1(\Omega)$ se da en el Teorema 3.2.1. De de las propiedades de monotonicidad de θ' , el esquema MDGLS se puede escribir como:

$$\widehat{L}_{k+1}^{n}(x) := \begin{cases} \tau \theta' (b_{k+1}^{n}(x)), & \text{si } b_{k+1}^{n}(x) < \widehat{z}, \\ \tau L_{\theta}, & \text{si } a_{k+1}^{n}(x) \le \widehat{z} \le b_{k+1}^{n}(x), \\ \tau \theta' (a_{k+1}^{n}(x)), & \text{si } a_{k+1}^{n}(x) > \widehat{z}. \end{cases}$$

De manera similar se puede definir el esquema MGLS por:

$$\widehat{L}_{k+1}^{n}(x) := \begin{cases} \frac{\theta\left(b_{k+1}^{n}(x) + d_{k+1}^{n}\right) - \theta\left(b_{k+1}^{n}(x)\right)}{d_{k+1}^{n}}, & \text{si } b_{k+1}^{n}(x) + d_{k+1}^{n} < \widehat{z}, \\ L_{\theta}, & \text{si } a_{k}^{k+1}(x) - d_{k+1}^{n} \le \widehat{z} \le b_{k+1}^{n}(x) + d_{k+1}^{n}, \\ \frac{\theta\left(a_{k+1}^{n}(x)\right) - \theta\left(a_{k+1}^{n}(x) - d_{k+1}^{n}\right)}{d_{k+1}^{n}}, & \text{si } a_{k+1}^{n}(x) - d_{k+1}^{n} > \widehat{z}. \end{cases}$$

3.4 Experimentos Numéricos

Mostraremos la eficiencia y robustez de los nuevos esquemas propuestos a través de tres ejemplo numéricos de referencia dados en la literatura. Presentamos resultados numéricos en dos y tres dimensiones. Para la discretización del tiempo, utilizamos el método de Euler semi-implícito. Para el espacio, elegimos elementos finitos lineales de Galerkin en mallas triangulares. Resumimos los resultados en la Tabla 3.1, 3.2, 3.3, 3.4 y 3.5, donde comparamos el *L-scheme* y el método de *Newton* con los nuevos esquemas de linealización. Todos los cálculos se realizaron en un portátil HP, con un procesador Intel Core i5-7400, 8 GB de RAM y el software Freefem ++ v4.6.

3.4.1 Ejemplo 1

Este ejemplo fue usado en [15] para comparar el *L*-scheme con una solución exacta. Extendemos el ejemplo a tres dimensiones con una nueva solución analítica. Consideremos la siguiente ecuación diferencial parcial no lineal elíptica en $H^1(\Omega)$, donde $\Omega = (0, 1)^3 \subset \mathbb{R}^3$:

$$\begin{aligned} \theta(\psi) - \Delta \Psi &= f \quad \text{en } \Omega, \\ \Psi &= w \qquad \text{sobre } \partial \Omega. \end{aligned}$$
 (3.18)

La función no lineal $\theta \in C^1(\Omega)$ está definida por:

$$\theta(\Psi) = \begin{cases} \arctan(\Psi), & \text{si } \Psi < 1, \\ \frac{\pi}{4}, & \text{si } \Psi \ge 1. \end{cases}$$

Su derivada está dada por:

$$\boldsymbol{\theta}'(\boldsymbol{\Psi}) = \begin{cases} \frac{1}{1+\Psi^2}, & \mathrm{si} \quad \Psi < 1, \\ 0, & \mathrm{si} \quad \Psi \ge 1. \end{cases}$$

Notamos que $0 \leq \theta'(\Psi) \leq 1$ para todo $\Psi \in \mathbb{R}$. El lado derecho f se elige de tal manera que la solución exacta de la ecuación diferencial (3.18) sea

$$\Psi(x, y, z) = w(x, y, z) = x^3 + y^3 - z^2 + x + y + \sin(\pi x)\sin(\pi y)\sin(\pi z).$$

Iteraciones	Newton	L-scheme	DGLS	GLS	MNS	MDGLS	MGLS				
h = 0.2165											
1	0.0395485311	0.0442408850	0.0395498562	0.0429041402	0.0395498562	0.0442408850	0.0442408850				
2	0.0006949759	0.0018067383	0.0011237906	0.0011691134	0.0006950585	0.0012508027	0.0016297909				
3	0.0006949658	0.0007153533	0.0006994734	0.0006998617	0.0006949658	0.0006960574	0.0006975815				
4	0.0006949658	0.0006955932	0.0006950395	0.0006950457	0.0006949658	0.0006949683	0.0006949723				
5	0.0006949658	0.0006949863	0.0006949671	0.0006949672	0.0006949658	0.0006949658	0.0006949658				
6	0.0006949658	0.0006949665	0.0006949659	0.0006949659	0.0006949658	0.0006949658	0.0006949658				
7	0.0006949658	0.0006949659	0.0006949658	0.0006949658	0.0006949658	0.0006949658	0.0006949658				
8	0.0006949658	0.0006949658	0.0006949658	0.0006949658	0.0006949658	0.0006949658	0.0006949658				
h = 0.0866											
1	0.0392926629	0.0439885584	0.0392934598	0.0426491121	0.0392934598	0.0439885584	0.0439885584				
2	0.0000419001	0.0014111014	0.0006480176	0.0007023323	0.0000419376	0.0008026772	0.0012192704				
3	0.0000418939	0.0000674276	0.0000450716	0.0000454444	0.0000418939	0.0000425269	0.0000436685				
4	0.0000418939	0.0000421967	0.0000419269	0.0000419297	0.0000418939	0.0000418950	0.0000418968				
5	0.0000418939	0.0000419030	0.0000418944	0.0000418945	0.0000418939	0.00004189 <mark>39</mark>	0.0000418939				
6	0.0000418939	0.0000418942	0.00004189 <mark>39</mark>	0.00004189 <mark>39</mark>	0.0000418939	0.0000418939	0.0000418939				
7	0.0000418939	0.00004189 <mark>39</mark>	0.0000418939	0.0000418939	0.0000418939	0.0000418939	0.0000418939				
8	0.0000418939	0.0000418939	0.0000418939	0.0000418939	0.0000418939	0.0000418939	0.0000418939				

Tabla 3.1: $\|\Psi_n - \Psi\|$ para el Ejemplo 1.

El valor del parámetro L que usamos en el L-scheme es el sugerido en [15], L = 1 y para DGLS y GLS, seleccionamos r = 10. Comenzamos las iteraciones con la solución inicial $\Psi_0(x, y, z) = -|x - 1.7|^{0.4} - 0.7|y - 1.7|^{1.6} + |z - 1.7|^{0.4}$.

Estudiamos dos tamaños de malla, h = 0.2165 y h = 0.0866. Comparamos las soluciones numéricas de los esquemas estudiados con la solución exacta, calculando el error y el gradiente del error en la norma $L^2(\Omega)$ en cada iteración. Los resultados se muestran en la Tabla 3.1 y 3.2 con diez dígitos de precisión.

En todos los esquemas, notamos que los errores disminuyen rápidamente con cada iteración del método de linealización, para luego alcanzar un valor constante Esto se debe a que al inicio del proceso iterativo el error de linealización es dominante, pero luego el error de discretización es el dominante y la parte constante en las tablas presentadas significa que el error de discretización se ha alcanzado, es decir, el error de la linealizacion está por abajo de los 10 dígitos decimales considerados. Es importante resaltar que este fenómeno ya fue observado en [15].

El esquema de Newton y el MNS entregan los mejores resultados, ya que en la tercera iteración se alcanzan el error de discretización para los dos errores analizados y para los dos tamaños de malla utilizados. El MDGLS y MGLS requieren cinco iteraciones para alcanzar el error de discretización en todos los casos. Los peores resultados se obtienen con el *L-scheme* porque requiere entre 7 y 8 iteraciones para alcanzar el error de discretización. Finalmente, DGLS y GLS necesitan una iteración menos que el método *L-scheme*.

Iteraciones	Newton	L-scheme	DGLS	GLS	MNS	MDGLS	MGLS				
h = 0.2165											
1	0.2243161693	0.2533833827	0.2243161744	0.2435031098	0.2243161744	.2243161744 0.2533833827					
2	0.0083799759	0.0120426968	0.0094677916	0.0096130623	0.0083801827	0.0099480237	0.0113066039				
3	0.0083799558	0.0084143298	0.0083872628	0.0083878966	0.0083799558	0.0083817090	0.0083841336				
4	0.0083799558	0.0083809627	0.0083800736	0.0083800836	0.0083799558	0.0083799603	0.0083799671				
5	0.0083799558	0.0083799885	0.0083799577	0.0083799579	0.0083799558	0.0083799558	0.0083799558				
6	0.0083799558	0.0083799 <mark>569</mark>	0.00837995 <mark>58</mark>	0.00837995 <mark>58</mark>	0.0083799558	0.0083799558	0.0083799558				
7	0.0083799558	0.00837995 <mark>58</mark>	0.0083799558	0.0083799558	0.0083799558	0.0083799558	0.0083799558				
8	0.0083799558	0.0083799558	0.0083799558	0.0083799558	0.0083799558	0.0083799558	0.0083799558				
h = 0.0866											
1	0.2233708172	0.2525153998	0.2233708224	0.2425718284	0.2233708224	0.2525153998	0.2525153998				
2	0.0005982444	0.0077362351	0.0035754006	0.0038693529	0.0005983370	0.0045490965	0.0066797904				
3	0.0005982302	0.0006583948	0.0006036515	0.0006043607	0.0005982302	0.0005994185	0.0006017325				
4	0.0005982302	0.0005986630	0.0005982753	0.0005982792	0.0005982302	0.0005982319	0.0005982344				
5	0.0005982302	0.0005982425	0.0005982 <mark>309</mark>	0.0005982 <mark>309</mark>	0.0005982302	0.00059823 <mark>02</mark>	0.00059823 <mark>02</mark>				
6	0.0005982302	0.0005982 <mark>306</mark>	0.000598230 <mark>2</mark>	0.000598230 <mark>2</mark>	0.0005982302	0.0005982302	0.0005982302				
7	0.0005982302	0.000598230 <mark>2</mark>	0.0005982302	0.0005982302	0.0005982302	0.0005982302	0.0005982302				
8	0.0005982302	0.0005982302	0.0005982302	0.0005982302	0.0005982302	0.0005982302	0.0005982302				

Tabla 3.2: $\|\nabla \Psi_n - \nabla \Psi\|$ para el Ejemplo 1.

3.4.2 Ejemplo 2

Este ejemplo fue propuesto en [4], es un problema en dos dimensiones espaciales y fue utilizado para estudiar comparativamente algunos esquemas de linealización. En esta tesis usamos este ejemplo para comparar los nuevos esquemas de linealización propuestos en la sección anterior y seguiremos la misma metodología sugerida en [4]. Fijamos una condición de parada y calculamos el número de iteración y el tiempo de computo siguiendo los trabajos [2, 4, 14, 15], además calculamos el orden de convergencia y solo para los esquemas lineales la tasa de convergencia.

Este ejemplo trata de un problema de inyección y extracción de fluido en dos dimensiones espaciales (x, z), donde consideramos la coordenada z hacia arriba. El dominio Ω esta dividido en dos partes: la zona no saturada $\Omega_v = (0, 1) \times (-3/4, 0)$ y la zona saturada $\Omega_s = (0, 1) \times (-1, -3/4)$. De esta manera el dominio del problema $\Omega = \Omega_v \cup \Omega_s = (0, 1) \times (0, 1)$. Las condiciones de contorno son mixtas, una condición de Dirichlet no homogénea en $\Gamma_D = (0, 1) \times \{0\}$ y una condiciones de Neumann sin flujo en el complemento $\Gamma_N = \partial \Omega - \Gamma_D$

$$\psi = -3$$
 sobre $\Gamma_D := (0,1) \times \{0\}$ y $[-K(\psi)\nabla(\psi+z)] \cdot \vec{n} = 0$ sobre $\Gamma_N := \partial \Omega \setminus \Gamma_D$.

La condición inicial es discontinua en la transición desde la zona saturada hacia la zona vadosa y está dada por:

$$\psi(x, z, 0) = \psi_0(x.z) = \begin{cases} -3 & \text{si } (x, z) \in \Omega_v, \\ -z - \frac{3}{4} & \text{si } (x, z) \in \Omega_s. \end{cases}$$

Seleccionamos un término fuente que toma valores positivos y negativos dado por:

$$f(x,z,t) = \begin{cases} -0.006\cos(\frac{3}{4}\pi z)\sin(2\pi x) & \text{si } (x,z) \in \Omega_v, \\ 0 & \text{si } (x,z) \in \Omega_s. \end{cases}$$

Las funciones hidráulicas involucradas están dadas mediante el modelo de Van-Genuchten con los siguiente parámetros: $\alpha = 0.95$, n = 2.9, $\theta_s = 0.42$, $\theta_r = 0.026$, y $K_s = 0.12$. La elección de n > 2 implica la Lipschitz continuidad de θ y K. Analizamos las soluciones numéricas después del primer paso de tiempo como fue sugerido en [4]. Tomamos $h_t \in \{0.25, 1, 5\}$ y usamos varios tamaños de malla h. Es fácil determinar que $L_{\theta} = 0.25$, usamos este valor en los métodos *L*-scheme y *L*2scheme según lo establecido en 3.3.2. Para los esquemas *DGLS* y *GLS*, utilizamos r = 10. Para detener las iteraciones, adoptamos el criterio $\|\psi_{k+1}^{n+1} - \psi_{k+1}^n\|_{L^2(\Omega)} \leq 0.001$ siguiendo [2, 15]. Comenzamos las iteraciones con la solución obtenida en el paso tiempo anterior $\Psi_{k+1}^0 = \Psi_k$.

Los resultados son presentados en la Tabla 3.3, observamos que el método de Newton falla para los dos tamaños de malla h más pequeños en todos los pasos de tiempo. Este resultado concuerda con los resultados teóricos discutidos en 2.1.2, donde se concluye que la convergencia del método de Newton depende del tamaño de malla h. Todos los demás esquemas convergen para todos los pasos de tiempo y tamaños de malla. También notamos que el esquema MGLS presenta los mejores resultados de entre todos los esquema de primer orden y el MNS es el mejor esquema de los métodos de segundo orden mostrando un comportamiento similar al método de Newton.

El número de iteraciones se muestra en la Tabla 3.3 y en la Figura 3.2. Concluimos que los nuevos esquemas mejoran la convergencia del *L-scheme* y el *L2-scheme*. Por otro lado, los esquemas DGLS y GLS obtienen las mismas iteraciones aunque el DGLS usa θ' . El *L-scheme* y el *L2-scheme* producen los peores resultados en lo que respecta al número de iteraciones, de hecho el *L-scheme* hace hasta 3 veces más iteraciones que el MNS. Los esquemas MDGLS y MGLS obtienen un rendimiento aceptable, entre 6 y 7 iteraciones. Finalmente, los esquemas de *Newton* y MNS obtuvieron los mejores resultados con un número de iteraciones más bajo. Más aún, el esquema MNS mantiene un número de iteraciones pequeño para todos los pasos de tiempo y tamaños de malla analizados.

Se calcula el orden de convergencia de los esquemas y en el caso de los esquemas con orden lineal calculamos además la tasa de convergencia. Los resultados se pueden ver en la Tabla 3.3. Los experimentos numéricos no muestran una variación significativa cuando cambiamos el paso de tiempo o la malla espacial. Los esquemas de *Newton* y *MNS* tiene un orden de convergencia experimental muy cercano a dos, mientras los restantes esquemas tienen un orden lineal. Un análisis más detallado de los esquemas de orden lineal se puede obtener comparando sus tasas de convergencia. Las mejores tasas de convergencia se obtienen para los esquemas *MDGLS* y *MGLS*, mientras que la tasa más pobre corresponde al *L-scheme*, lo que demuestra que los nuevos esquemas

	Iteraciones						Tiempo de CPU				Convergencia	
h	0.304	0.160	0.075	0.033	0.019	0.304	0.160	0.075	0.033	0.019	Orden	Tasa
$h_t = 0.25$												
Newton	4	5	6	falla	falla	0.11	0.49	2.31	falla	falla	1.97	_
L-scheme	15	12	12	13	13	0.25	0.84	3.30	23.39	70.52	1	0.87
L2-scheme	11	13	8	8	8	0.22	1.15	2.40	15.09	46.42	1	0.74
DGLS	11	9	7	8	8	0.34	1.61	3.67	27.10	89.10	1	0.72
GLS	11	9	7	8	8	0.27	1.12	3.45	21.23	65.77	1	0.71
MNS	6	6	5	5	6	0.13	0.53	1.71	10.26	44.45	1.96	-
MDGLS	7	7	6	7	8	0.13	0.53	1.78	13.49	49.13	1	0.41
MGLS	7	7	7	8	8	0.15	0.57	2.12	16.74	50.30	1	0.42
$h_t = 1$												
Newton	4	5	6	falla	falla	0.09	0.50	2.32	falla	falla	1.96	_
L-scheme	14	15	16	16	16	0.23	1.08	4.41	26.87	86.34	1	0.87
L2-scheme	13	17	11	10	10	0.33	1.70	3.55	18.65	59.37	1	0.73
DGLS	10	9	10	10	10	0.32	1.31	5.25	32.14	107.99	1	0.72
GLS	10	9	10	10	10	0.25	1.00	4.33	25.62	84.37	1	0.73
MNS	5	5	5	5	6	0.10	0.46	1.74	10.89	40.00	1.94	_
MDGLS	6	6	6	7	7	0.11	0.46	1.81	12.70	41.08	1	0.41
MGLS	7	7	7	7	8	0.13	0.56	2.10	12.88	47.56	1	0.42
$h_t = 5$												
Newton	4	5	9	falla	falla	0.09	0.50	3.44	falla	falla	1.98	_
L-scheme	17	19	19	19	19	0.28	1.33	5.22	31.98	106.12	1	0.86
L2-scheme	13	16	14	12	12	0.27	1.84	5.59	20.62	77.86	1	0.74
DGLS	10	11	11	11	11	0.31	1.61	6.26	35.21	114.69	1	0.72
GLS	10	11	11	11	11	0.25	1.22	4.58	27.80	87.50	1	0.73
MNS	5	5	5	5	5	0.10	0.44	1.71	10.66	35.66	1.95	_
MDGLS	7	7	6	6	7	0.13	0.53	1.75	10.84	43.09	1	0.41
MGLS	7	7	7	7	7	0.13	0.56	2.15	13.57	41.42	1	0.42

Tabla 3.3: Número de iteraciones, tiempo de CPU y orden de convergencia para $h_t = 0.25, 1, 5$ y cinco tamaños de mallas h diferentes.



Figura 3.2: Número de iteraciones.



Figura 3.3: Tiempo de CPU.



Figura 3.4: Evolución de L_1^n para el esquema *MNS* con el paso de tiempo $h_t = 0.25$ y un tamaño de malla de h = 0.027.

propuestos en esta tesis son una mejor alternativa. Puesto que el L2-scheme de alguna manera mejora la aproximación de la derivada, este calcula menos iteraciones que el L-scheme, lo que se refleja en la rapidez del método.

Los buenos resultados que se obtienen con el esquema MNS se debe a la aproximación de θ' por L_{k+1}^n . En la Figura 3.4 podemos observar cómo la función L_{k+1}^n en color azul, se acerca a θ' en color rojo a medida que las iteraciones avanzan. Más aún en la cuarta iteración L_{k+1}^n ya ha alcanzado el valor de θ' , por lo que a partir de esta iteración el esquema corresponde al método de Newton.

Finalmente, nos interesa el tiempo de CPU de los nuevos esquemas, estos se comparan con los métodos *L-scheme*, *L2-scheme* y Newton. Los resultados se muestran en la Tabla 3.3 y en la Figura 3.3 solo para $h_t = 1$. Téngase en cuenta que en la Figura 3.3 se usa 1/h en lugar de *h* para una mejor visualización de los resultados. El tiempo de CPU aumenta cuando usamos tamaños de malla pequeños sin embargo no se exhibe mayor variabilidad cuando se cambia el paso de tiempo. Aún cuando los esquemas *DGLS* y *GLS* usan el mismo número de iteraciones, el *GLS* registra un tiempo de CPU ligeramente más bajo, esto se debe a que no usa la derivada y por lo tanto el costo de cada iteración es menor. El *DGLS* es el método lineal más lento siendo incluso más lento que el *L-scheme* en todos los casos. Los esquemas *MGLS* y *MDGLS* registran tiempos de CPU similares, sin embargo el tiempo de CPU del método *MDGLS*, es ligeramente más bajo y esto se debe a que cada iteración es más costosa, ya que necesita evaluar la función θ muchas más veces. En todo caso el tiempo de cómputo de estos dos métodos es significativamente más pequeño que el de los esquemas DGLS y GLS. El MNS es el esquema que muestra el mejor rendimiento en lo que a tiempo de CPU se refiere, aunque es importante rescatar que cuando se usa mallas muy gruesas, el método de Newton registra un tiempo de CPU ligeramente mejor que el MNS. Sin embargo en la iteración inmediatamente anterior a la falla del método de Newton, este tiene un aumento significativo en el tiempo de cálculo superando significativamente al esquema MNS.

3.4.3 Ejemplo 3

Este ejemplo ha sido usado por varios autores y para algunos detalles podemos referirnos a [3,22,26,27]. Fue propuesto por primera vez por Celia et al. en 1990 [3] para estudiar el método de linealización de *Picard*. Considera el caso de un suelo arenoso representado por el dominio espacial $\Omega = (0,2) \times (-40,0)$. Las funciones hidráulicas involucradas están dadas por un modelo Haverkamp [8], correspondiente a las ecuaciones (1.6) y (1.7). Los parámetros del modelo vienen dados por: $\theta_s = 0.287$, $\theta_r = 0.075$, $\alpha = 1.611 \times 10^6$, $\beta = 3.96$, $K_s = 9.44 \times 10^{-3}$, $A = 1.175 \times 10^6$, y $\gamma = 4.74$. Fijamos la condición de parada $\|\psi_{k+1}^{n+1} - \psi_{k+1}^n\|_{L^2(\Omega)} \leq 10^{-10}$ y calculamos el número de iteraciones, tiempo de CPU y orden de convergencia. En este ejemplo no comenzamos las iteraciones de los métodos de linealización con la solución en el tiempo anterior como es habitual, mas bien comenzamos todas las iteraciones con la solución inicial fija $\psi_{k+1}^0(x, z) = 1.02z - 20.7$. Analizamos las soluciones numéricas para los pasos de tiempo $h_t = 1$, 10 y los tamaños de malla h = 0.9428, 0.6248, 0.3195. Además estudiamos el problema usando dos condiciones iniciales $\Psi_0 = -61.5$ y $\Psi_0 = -40$.

Los resultados de convergencia se muestran en la Tabla 3.4 para la primera condición inicial y en la Tabla 3.5 para la segunda condición inicial. Observamos que todos los esquemas son convergentes, inclusive el esquema de *Newton*, para todos las pasos de tiempo, tamaños de malla y condiciones iniciales utilizadas. Además el tamaño de la malla no tiene un efecto significativo en el número de iteraciones de ninguno de los métodos de linealización estudiados. De todos los esquemas de orden lineal, el MGLS logra los mejores resultados, y además el MNS supera a todos los esquemas inclusive al método de *Newton*.

Analizando el número de iteraciones en los esquemas lineales, notamos que el *L*scheme, con la condición inicial $\Psi_0 = -61.5$, produce los peores resultados con 114 iteraciones, mientras el *L2-scheme*, tiene un rendimiento aceptable de 45 iteraciones. Sin embargo, si cambiamos la condición inicial a $\Psi_0 = -40$ los resultados más pobres los obtiene el *L2-scheme* con aproximadamente 70 iteraciones, mientras que el *L-scheme* tiene resultados aceptables, alrededor de 18 iteraciones.

Los esquemas DGLS y GLS usan entre 44 y 45 iteraciones con la condición inicial $\Psi_0 = -61.5$, mientras que si usamos $\Psi_0 = -40$ los dos esquemas mejoran su rendimiento. Más aún, el esquema DGLS es mejor que el GLS ya que el primero registra aproximadamente 8 iteraciones y el otro entre 16 y 18 iteraciones. Los es-

Tabla 3.4: Número de iteraciones, tiempo de CPU y orden de convergencia para los tamaños de malla 0.9428, 0.6248 y 0.3195, con pasos de tiempo $h_t = 1, 10$, usando la condición inicial $\Psi_0 = -61.5$.

	Tier	mpo de O	Convergencia						
h	0.9428	0.6248	0.3195	0.9428	0.6248	0.3195	Orden	Tasa	
		ŀ	$n_t = 1$						
Newton	7	7	7	0.96	2.84	10.59	1.85	_	
L-scheme	114	114	114	10.25	30.46	133.78	1	0.76	
L2-scheme	45	45	45	4.87	12.36	47.71	1	0.53	
DGLS	44	44	44	8.14	23.14	88.34	1	0.53	
GLS	45	44	44	6.54	19.91	78	1	0.53	
MNS	6	6	6	1.19	4.78	17.28	1.93	—	
MDGLS	8	8	9	1.81	7.03	28.96	1	0.12	
MGLS	9	9	9	2.08	8.68	$28,\!96$	1	0.12	
$h_t = 10$									
Newton	7	7	7	0.95	2.87	10.52	1.81	_	
L-scheme	114	114	114	10.19	30.17	114.52	1	0.76	
L2-scheme	45	45	45	4.01	11.98	46.87	1	0.53	
DGLS	44	44	44	7.84	23.19	88.07	1	0.53	
GLS	44	44	44	8.49	19.24	73.68	1	0.53	
MNS	6	6	6	1.61	4.52	17.78	1.90	_	
MDGLS	9	9	9	2.02	7.94	29.68	1	0.12	
MGLS	9	9	9	2.05	8.42	31.17	1	0.12	
	Iteraciones			Tier	Tiempo de CPU			Convergencia	
-----------	-------------	--------	-----------	--------	---------------	--------	-------	--------------	--
h	0.9428	0.6248	0.3195	0.9428	0.6248	0.3195	Orden	Tasa	
		ŀ	$n_t = 1$						
Newton	8	8	8	1,24	$3,\!39$	12,66	1.75	_	
L-scheme	17	18	19	1.70	5.28	20.66	1	0.15	
L2-scheme	74	74	74	6.66	21.21	78.90	1	0.67	
DGLS	8	9	9	1.60	5.13	18.68	1	0.01	
GLS	16	17	18	2.34	7.80	29.97	1	0.16	
MNS	5	5	5	1.43	4.35	14.84	1.92	—	
MDGLS	11	11	11	2.26	9.75	35.80	1	0.13	
MGLS	11	11	11	2,53	10.29	37.18	1	0.13	
		h	t = 10						
Newton	7	7	7	1.02	3.40	11.12	1.80	_	
L-scheme	19	19	19	1.84	5.45	20.56	1	0.23	
L2-scheme	71	71	71	7.86	19.98	79.58	1	0.66	
DGLS	8	9	9	1.55	5.00	18.66	1	0.02	
GLS	18	18	18	2.74	8.55	31.17	1	0.16	
MNS	5	5	5	1.36	3.86	14.80	1.85	—	
MDGLS	11	11	11	2.50	9.74	35.85	1	0.13	
MGLS	11	11	11	2.55	10.49	37.18	1	0.13	

Tabla 3.5: Número de iteraciones, tiempo de CPU y orden de convergencia para los tamaños de malla de 0.9428, 0.6248 y 0.3195, con pasos de tiempo $h_t = 1, 10$, usando la condición inicial $\Psi_0 = -40.0$

quemas MDGLS y MGLS tiene buenos resultados con las dos condiciones iniciales, con la primera usa alrededor de 9 iteraciones y con la otra 11.

En el caso de los esquemas de orden superior, el mejor resultado se obtiene con el esquema MNS seguido muy de cerca por el método de Newton, para las dos condiciones iniciales utilizadas. Es así que el MNS usa entre 5 y 6 iteraciones y el Newton entre 7 y 8.

El orden de convergencia y las tasas de convergencia para los esquemas lineales no exhiben mayor variabilidad cuando cambiamos el paso de tiempo, sin embargo las tasas de convergencia sí presentan cambios significativos cuando cambiamos la condición inicial. Los esquemas de *Newton* y *MNS* muestran un orden de convergencia experimental similar y cercano a dos, para todos los pasos de tiempo y condiciones iniciales usadas, los demás esquemas tienen un orden lineal.

La mejor tasa de convergencia se obtiene para el esquema DGLS con la segunda condición inicial, pero este esquema no consigue una buena tasa de convergencia cuando usamos la primera condición inicial. De hecho en este caso los esquemas



Figura 3.5: Tiempo de CPU para $h_t = 1$ con la condición inicial $\Psi_0 = -61.5$.

MDGLS y *MGLS* exhiben la mejor tasa de convergencia. La peor tasa de convergencia corresponde al *L-scheme* con la primera condición inicial. Finalmente, con la segunda condición inicial su tasa de convergencia mejora significativamente, de hecho en este caso la peor tasa de convergencia es para el *L2-scheme*.

Las tasas de convergencia con la condición inicial $\Psi_0 = -61.5$ se presentan en la Tabla 3.4, los esquemas *MDGLS* y *MGLS* exhiben la mejor tasa con un valor de 0.12 seguido de los es esquemas *DGLS*, *GLS* y *L2-scheme* que registran una tasa de 0.53. La peor tasa de convergencia es la del *L-scheme* con 0.76.

Los resultados con la condición inicial $\Psi_0 = -40$ se tienen en la Tabla 3.5, el método DGLS tiene la mejor tasa de convergencia con valores de 0.01 y 0.02 respectivamente para los dos pasos de tiempo utilizados, en segundo lugar se ubican los esquemas MDGLS y MGLS con una tasa de 0.13, seguido del GLS con 0.16, luego el *L-scheme* con tasas de 0.15 y 0.23 respectivamente para los dos pasos de tiempo utilizados. Finalmente la peor tasa de convergencia es 0.67 para el *L2-scheme*.

Es importante resaltar que los esquemas MGLS y MDGLS tienen un buen comportamiento mostrando tasas de convergencia bajas con cualquiera de las dos condiciones iniciales usadas, en contraste con el *L-scheme* y *L2-scheme* que muestran cierta inestabilidad, pues sus tasas de convergencia cambian considerablemente cuando cambiamos la condición inicial.

En la Tabla 3.4 y la Figura 3.5 se muestra los tiempos de CPU para la condición inicial más seca. El esquema más rápido es el de *Newton*, aún cuando utiliza más iteraciones que el esquema MNS, esto se debe al hecho que cada iteración del método MNS es más costosa, puesto que hace más evaluaciones de la derivada. Por otro lado, el *L-scheme* es el peor, ya que es casi 8 veces más lento que el MNS y 10 veces más lento que el esquema de *Newton*. El método más rápido de todos los que no usan la



Figura 3.6: Tiempo de CPU para $h_t = 1$ con la condición inicial $\Psi_0 = -40$.

derivada de θ es el esquema MGLS, seguido del L2-schemey finalmente el esquema GLS.

Con la condición inicial $\Psi_0 = -40$, los resultados se muestran en la Tabla 3.5 y en la Figura 3.6. Nuevamente no notamos variaciones importantes del tiempo de CPU cuando cambiamos el paso de tiempo. De los métodos que no usan θ' , el *L-scheme* muestra los mejores resultados, seguido del esquema *GLS*. Los mejores tiempos de cálculo corresponden a los métodos de *Newton* y *MNS*. El esquema *DGLS* también muestra buenos resultados y es sólo un 20 % más lento que el esquema *MNS*. Por otro lado, *L2-scheme* es el método más lento es aproximadamente 5 veces más lento que el *MNS*, lo que muestra un resultado muy pobre.

Observación 3.4.1. Notemos que el esquema MNS tiene un mejor desempeño que el método de Newton en el Ejemplo 3. Esto no debería sorprendernos, pues en [4] ya fue reportado un esquema que al igual que el MNS inicia las iteraciones con un método de orden lineal, para luego cambiar a Newton, mostrando un mejor desempeño que el método de Newton. El esquema presentado en [4] es el L-esquema/Newton y es un esquema mixto que consiste en comenzar las iteraciones con el L-scheme para luego cambiar al método de Newton. Sus resultados pueden ser consultados en [4, Fig 1]. Pensamos que esto podría obedecer a un pobre desempeño del esquema de Newton en las primeras iteraciones, cuando la solución inicial está muy lejos de la solución del problema, es decir, en las primeras iteraciones el método de Newton, podría ser más lento que los métodos de orden lineal. Además se debe tener en cuenta que no hemos elegido la solución en el tiempo anterior como la solución inicial, lo que de alguna manera podría haber contribuido para que la solución inicial este lejos de la solución del problema.

3.5 Conclusiones

Este capítulo de la tesis trata de una generalización del L-scheme, que es un esquema de primer orden globalmente convergente. Al estudiar el rendimiento de los esquemas numéricos L-scheme y L2-scheme, notamos una fuerte dependencia de la condición inicial, ya que para $\Psi_0 = -61.5, \frac{1}{2}L_{\theta}$ es mejor que L_{θ} , mientras que para la condición inicial $\Psi_0 = -40 L_{\theta}$ funciona mejor. Con base en este hecho, se propuso un marco teórico que permitió generalizar estos esquemas, concretamente el Teorema 3.2.1 que permite obtener otros esquemas de primer orden globalmente convergentes. En base a esta teoría se pueden desarrollar muchos esquemas pero hemos seleccionamos cuatro métodos de primer orden (GLS, DGLS, MGLS, MDGLS) y uno de segundo orden el MNS para ilustrar el marco teórico. Proponemos dos esquemas basados en una aproximación del término no lineal que no usan su derivada y los otros dos esquemas, el DGLS y MDGLS, que usan derivadas para aproximar alguna parte de la no linealidad. El objetivo principal de estos dos últimos esquemas es medir el efecto de la derivada en la convergencia de los esquemas. De manera similar, proponemos el MNS que es un esquema con convergencia similar al método de Newton pero mucho más robusto por lo constituye una valiosa es una alternativa.

El primer ejemplo numérico de referencias presentado en este capítulo tiene la finalidad de comparar los nuevos esquemas propuestos con el L-scheme y el método de Newton usando la solución exacta de la ecuación diferencial no lineal. En este ejemplo analizamos el número de iteraciones que necesita el método para alcanzar el error de discretización, esto permite sacar las primeras conclusiones sobre los nuevos esquemas propuestos. En la Tabla 3.1 y 3.2 se aprecia un pobre rendimiento del L-scheme, los esquemas de Newton y MNS tienen los mejores resultados, alcanzando rápidamente el error de discretización, así mismo los esquemas MDGLS y MGLS tienen un rendimiento similar superando de manera significativa al L-scheme, los métodos DGLS y GLS muestran una convergencia parecida y también superan al L-scheme.

El segundo ejemplo numérico de referencia es un ejemplo donde el esquema de Newton falla, esto permite analizar la robustez de los esquemas. Se concluye que los nuevos esquemas son robustos y mejoraron la convergencia, siendo más rápidos que el L-scheme y el L2-scheme, entre dos y cuatro veces más rápido, mostrando un tiempo de CPU en general más bajo. Además notamos que en la mayoría de los casos obtenemos un número de iteración más bajo, manteniendo la propiedad de convergencia global, lo que es importante para los problemas singulares donde falla el método de Newton, como se muestra en la Tabla 3.3 cuando h disminuye. Observamos además, que el esquema MNS no es notablemente mejor que el método MGLS, registrando solo unas pocas iteraciones más y un tiempo de cálculo similar. Los esquemas lineales que usan la derivada de θ , obtienen resultados muy similares a las contrapartes que no la usan y todos los esquemas propuestos muestran un número de iteración estable a medida que h disminuye. También concluimos para este ejemplo de referencia, que el método MGLS se comporta de manera similar a los métodos de segundo orden mostrando pocas iteraciones y tiempos de CPU pequeños.

El tercer ejemplo de referencia muestra un comportamiento muy diferente de los métodos L-scheme y L2-scheme. En el primer caso con $\Psi_0 = -61.5$, el L-esquema requirió tres veces más iteraciones que el L2-scheme. También notamos que los métodos GLS y DGLS se comportan de manera similar al L2-scheme, mientras el MGLS hace una quinta parte de las iteraciones de estos esquemas lo que mejora significativamente el tiempo de CPU, aunque cada iteración es más costosa que una iteración del esquema GLS. Aún cuando el método MNS tiene una iteración menos que el método de Newton, este requiere más tiempo de computo. En el segundo caso cuando $\Psi_0 = -40$, el L2-scheme tiene aproximadamente cuatro veces más iteraciones que el esquema Lscheme. El método DGLS muestra una gran mejora con respecto al esquema GLS, mostrando un rendimiento parecido al método de Newton. Además se pudo observar que el MNS obtiene menos iteraciones que el método de Newton, sin embargo registra un tiempo de CPU ligeramente mayor. En el caso de los esquemas MGLS y MDGLS, estos tienen un rendimiento similar aunque no mejor que el método MNS. Concluimos que entre todos los métodos de orden lineal el MGLS es el que tiene el mejor rendimiento, sin embargo el esquema MNS es el mejor de todos los métodos propuestos superando incluso al método de Newton. El esquema MNS tiene el mismo orden de convergencia que el esquema de Newton, pero es un método más robusto por lo que se convierte en una valiosa alternativa al esquema de Newton.

Finalmente, concluimos que el método MGLS funciona mejor que los esquemas Lscheme y L2-scheme, para todos los casos, de hecho este es el mejor esquema robusto con orden lineal de todos los métodos que hemos estudiado durante este capítulo. La tasa de convergencia de los nuevos métodos es pequeña y el rendimiento es muy bueno más aún en algunos casos puede ser tan bueno como el método de Newton. De todos los métodos que usan la derivada de θ , el MNS es el mejor esquema.

Estamos interesados en obtener esquemas globalmente convergentes y de segundo orden o con convergencia superlineal, pero siempre conservando la propiedad de convergencia global. Estamos trabajando en esta dirección y el próximo capítulo trata sobre esta temática.

Capítulo 4 NUEVOS ESQUEMAS DE ORDEN SUPERLINEAL

El objetivo de este capítulo es presentar tres nuevos esquemas de linealización eficientes y robustos con orden de convergencia superlineal para (2.5). Demostramos rigurosamente su convergencia y presentamos experimentos numéricos que confirman los hallazgos teóricos. Usamos la Ecuación de Richards en su forma mixta (1.8) y restringimos las formulaciones y el análisis a condiciones de frontera de Dirichlet homogéneas solo por simplicidad, sin embargo la extensión a condiciones de frontera más generales es sencilla, de hecho los ejemplos numéricos presentados en este capítulo consideren otras condiciones de borde. La discretización del tiempo se basa en el método de Euler semi-implícito y el espacio se discretiza mediante el método elementos finitos de Galerkin.

Proponemos tres nuevos esquemas de linealización para el problema totalmente discreto (2.6): L-Newton, Type-Secant y L-Secant. El primer método es una combinación eficiente de los métodos de Newton y L-scheme, esto da como resultado un esquema con orden de convergencia superlineal y globalmante convergente. El segundo esquema está inspirado en el método de la secante y tiene un orden superlineal, no usa las derivadas de las funciones involucradas y aún cuando solo es localmente convergente es más robusto que el esquema de Newton. El último método combina el segundo método con el L-scheme, consiguiendo un método con convergencia global, orden superlineal, y que no usa las derivadas de las funciones involucradas.

En el transcurso del capítulo demostramos formalmente la convergencia de las iteraciones de cada uno de los nuevos esquemas propuestos en la norma $H^1(\Omega)$. Los métodos de linealización se prueban en ejemplos numéricos ilustrativos en dos y tres dimensiones espaciales, donde se muestra la eficiencia y robustez de los esquemas propuestos.

El capítulo está organizado de la siguiente manera. En los tres siguientes apartados se proponen y discuten los tres nuevos esquemas iterativos, así como también probamos rigurosamente su convergencia y hacemos varias observaciones. La penúltima sección se refiere a los experimentos numéricos, aquí comparamos los esquemas propuestos con el *L-scheme* y el método *Newton*. El capítulo concluye con algunos comentarios finales. Asumimos (M1)-(M6) de las hipótesis 2.1.2 en el problema (2.6). Estas hipótesis fueron utilizadas en el apartado 2.1.2 para estudiar el método de *Newton* y su convergencia.

4.1 Esquema L-Newton

En todo lo que sigue y con la finalidad de simplificar la notación, omitiremos el superíndice h de las ecuaciones. Presentamos un esquema numérico de linealización para el problema completamente discreto (2.6) al que llamaremos **L-Newton**. El esquema propuesto converge independientemente de solución inicial escogida para comenzar las iteraciones del esquema, tiene un orden de convergencia de $1 + \gamma$ y no es necesario tomar ninguna restricción sobre el paso de tiempo. Para $k \in \{0, 1, ..., N\}$ fijo:

$$\begin{array}{l} \left(\text{Dados } \Psi_k, \Psi_{k+1}^n \in V_h, \text{ hallar } \Psi_{k+1}^{n+1} \in V_h \\ \left(1 - \lambda_{k+1}^n \right) L \left\langle \Psi_{k+1}^{n+1} - \Psi_{k+1}^n, v \right\rangle + \lambda_{k+1}^n \left\langle \theta'(\Psi_{k+1}^n)(\Psi_{k+1}^{n+1} - \Psi_{k+1}^n), v \right\rangle \\ &+ h_t \left\langle K(\Psi_k) \nabla(\Psi_{k+1}^{n+1} + z), \nabla v \right\rangle \\ &= \left\langle \theta_k - \theta_{k+1}^n + h_t S_{k+1}, v \right\rangle \quad \text{para todo } v \in V_h, \end{array}$$

$$\tag{4.1}$$

donde L es una constante y $\lambda_{k+1}^n \in [0,1]$ es una sucesión en \mathbb{R} que converge a 1. Si $\lambda_{k+1}^n = 1$ para todo n = 0, 1, ..., entonces el esquema es el método de Newton, ver sección 2.1.2. Por el contrario, si $\lambda_{k+1}^n = 0$ para todo n = 0, 1, ..., entonces el esquema corresponde al *L*-scheme, ver en la sección 2.3.2. A continuación analizaremos rigurosamente la convergencia de esquema (4.1). Los siguientes supuestos sobre los parámetros son necesarios para probar la convergencia:

Hipótesis 4.1.1. Escogemos el parámetro λ_{k+1}^n de la siguiente manera

(M14) Para todo k fijo, la sucesión $\{\lambda_{k+1}^n\}_{n\in\mathbb{N}}$ está dada por: $\lambda_{k+1}^0 = 0$ y para $n = 1, 2, \dots$

donde μ_{k+1}^n es una sucesión tal que:

$$\mu_{k+1}^n < A_{k+1}^n = \frac{(1+\gamma)h_t K_r h^{\frac{d\gamma}{2}}}{C_p^2 C_\theta \sqrt{C} \|\Psi_{k+1}^n - \Psi_{k+1}\|^{\gamma}},$$

 C_p es la constante de la desigualdad de Poincaré, h es el tamaño de la malla espacial y C es la constante positiva introducida en el Lema (2.1.2).

El principal resultado sobre la convergencia del método L-Newton es el siguiente teorema:

Teorema 4.1.1. Sea $k \in \{1, ..., N\}$ fijo. Suponga que se satisfacen las hipótesis (M1), (M2) y (M3). Elegimos $L \geq \frac{1}{2}L_{\theta}$ y λ_{k+1}^n como en la hipótesis (M14). Entonces la sucesión de la soluciones del problema (4.1) $\{\Psi_{k+1}^n\}_{n>0}$ converge a Ψ_{k+1} solución del problema (2.6) en la norma de $H^1(\Omega)$.

Demostración. Restando (2.6) de (4.1), tomando $v = e_{k+1}^{n+1} := \Psi_{k+1}^{n+1} - \Psi_{k+1} \in V_h$ y haciendo algunas operaciones algebraicas se consigue:

$$(1 - \lambda_{k+1}^n) L \left\langle e_{k+1}^{n+1} - e_{k+1}^n, e_{k+1}^{n+1} \right\rangle + \lambda_{k+1}^n \left\langle \theta'(\Psi_{k+1}^n) (e_{k+1}^{n+1} - e_{k+1}^n), e_{k+1}^{n+1} \right\rangle + h_t \left\langle K(\Psi_k) \nabla e_{k+1}^{n+1}, \nabla e_{k+1}^{n+1} \right\rangle = (1 - \lambda_{k+1}^n) \left\langle \theta_k - \theta_{k+1}^n, e_{k+1}^{n+1} \right\rangle + \lambda_{k+1}^n \left\langle \theta_k - \theta_{k+1}^n, e_{k+1}^{n+1} \right\rangle.$$

Usando ahora la hipótesis (M3) se obtiene

$$(1 - \lambda_{k+1}^{n})L \|e_{k+1}^{n+1}\|^{2} + \lambda_{k+1}^{n} \left\langle \theta'(\Psi_{k+1}^{n})e_{k+1}^{n+1}, e_{k+1}^{n+1} \right\rangle + h_{t}K_{r} \|\nabla e_{k+1}^{n+1}\|^{2}$$

$$\leq (1 - \lambda_{k+1}^{n}) \left\langle \theta_{k+1} - \theta_{k+1}^{n} - L(\Psi_{k+1} - \Psi_{k+1}^{n}), e_{k+1}^{n+1} \right\rangle$$

$$+ \lambda_{k+1}^{n} \left\langle \theta_{k+1} - \theta_{k+1}^{n} - \theta'(\Psi_{k+1}^{n})(\Psi_{k+1} - \Psi_{k+1}^{n}), e_{k+1}^{n+1} \right\rangle.$$

Sea $I_{k+1}^n := \{ z = \alpha \Psi_{k+1}^n + (1-\alpha) \Psi_{k+1}, \ \alpha \in [0,1] \} \subset \mathbb{R}, \text{ obtenemos}$

$$(1 - \lambda_{k+1}^n) L \|e_{k+1}^{n+1}\|^2 + h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 \leq (1 - \lambda_{k+1}^n) \left\langle \int_{I_{k+1}^n} |\theta'(t) - L| dt, |e_{k+1}^{n+1}| \right\rangle$$

$$+ \lambda_{k+1}^n \left\langle \int_{I_{k+1}^n} |\theta'(t) - \theta'(\Psi_{k+1}^n)| dt, |e_{k+1}^{n+1}| \right\rangle.$$

Usando (M1) y (M2) en los términos de la derecha tenemos

$$(1 - \lambda_{k+1}^{n})L \|e_{k+1}^{n+1}\|^{2} + h_{t}K_{r} \|\nabla e_{k+1}^{n+1}\|^{2} \leq (1 - \lambda_{k+1}^{n})L \left\langle |e_{k+1}^{n}|, |e_{k+1}^{n+1}| \right\rangle + \lambda_{k+1}^{n} \frac{C_{\theta}}{1 + \gamma} \left\langle |e_{k+1}^{n}|^{\gamma+1}, |e_{k+1}^{n+1}| \right\rangle.$$
(4.2)

Usando las desigualdades de Poincaré, Cauchy-Schwarz, Young y la estimación inversa dada en el Lema 2.1.2 se obtiene

$$\begin{split} \left[(1 - \lambda_{k+1}^n)L + \frac{h_t K_r}{C_p^2} \right] \|e_{k+1}^{n+1}\|^2 + &\leq (1 - \lambda_{k+1}^n)L \|e_{k+1}^n\|^2 \\ &+ (\lambda_{k+1}^n)^2 \frac{C_p^2 C_\theta^2 C}{K_r (1 + \gamma)^2 h_t h^{\gamma d}} \|e_{k+1}^n\|^{2+2\gamma}. \end{split}$$

Después de algunas simplificaciones, la desigualdad anterior se convierte en

$$\|e_{k+1}^{n+1}\| \le \beta_{k+1}^n \|e_{k+1}^n\|, \tag{4.3}$$

donde

$$\beta_{k+1}^{n} = \sqrt{\frac{C_{p}^{2}\left[(\lambda_{k+1}^{n})^{2}C_{p}^{2}C_{\theta}C\|e_{k+1}^{n}\|^{2\gamma} + (1-\lambda_{k+1}^{n})(1+\gamma)^{2}Lh^{\gamma d}h_{t}K_{r}\right]}{(1+\gamma)^{2}h^{\gamma d}h_{t}K_{r}\left[(1-\lambda_{k+1}^{n})C_{p}^{2}L + h_{t}K_{r}\right]}}$$

Usando la hipótesis (M14) es fácil demostrar que $\beta_{k+1}^n < 1$, lo que demuestra la convergencia en el espacio $L^2(\Omega)$. De la ecuación (4.2) y (4.3) se deduce que:

$$\|\nabla e_{k+1}^{n+1}\| \le \sqrt{\frac{\lambda_{k+1}^n C_\theta \sqrt{C}}{(1+\gamma)h^{\frac{\gamma d}{2}} h_t K_r}} \|e_{k+1}^n\|^{1+\frac{\gamma}{2}},$$

lo que finalmente demuestra la convergencia en el espacio $H^1(\Omega)$.

Observación 4.1.1. Puesto que $||e_{k+1}^n||$ converge a cero, es evidente que después de un número finito de iteraciones r, $\lambda_{k+1}^n = 1$ para todo n > r y de la desigualdad (4.2) deducimos que el esquema L-Newton tiene un orden de convergencia $\gamma + 1$. La convergencia puede volverse incluso cuadrática cuando θ' es Lipschitz continua. Si el método de Newton converge, tiene exactamente el mismo orden de convergencia que el esquema L-Newton. Observe que en las primeras iteraciones puede considerarse como un esquema cuasi-Newton, donde la función $\theta'(z)$ se aproxima por la función $L_{k+1}^n(z) = (1 - \lambda_{k+1}^n)L + \lambda_{k+1}^n \theta'(z)$ en la enésima iteración, esta aproximación se muestra en la Figura 4.3.

Observación 4.1.2. En la expressión A_{k+1}^n , $||e_{k+1}^n||$ se puede estimar usando la desigualdad (4.3) y en la primera iteración, $||e_{k+1}^0||$ se puede calcular usando la estimación demostrada en la Proposición 3.5 de [31]. Esta estimación ya fue usada para analizar algunos esquemas de linealización para la ecuación de Richard en [41].

Observación 4.1.3. La convergencia del esquema L-Newton es global, porque es independiente de la elección de la solución inicial con la que comencemos las iteraciones del método y de los parámetros de discretización. Sin embargo será beneficioso si se inician las iteraciones con la solución obtenida en el tiempo anterior. Por el contrario, el esquema de Newton aplicado a la Ecuación de Richards no es globalmente convergente, más aún, la convergencia del método de Newton depende de la elección de los parámetros de discretización, lo que ya ha sido reportado por algunos autores [4, 41, 43]. Siguiendo estos artículos, en la sección 2.1.2 se demostró el teorema de convergencia del método de Newton para la formulación (2.6). Según la estimación (2.8) de la sección 2.1.2, si fijamos el parámetro h, solo la elección de h_t suficientemente grande garantizará la convergencia. Alternativamente, si se fija h_t, el tamaño de la malla no puede ser demasiado pequeño. En caso de una divergencia en el esquema de Newton, refinar la malla no ayudará.

Observación 4.1.4. Si en lugar de la hipótesis (M3) tomamos simplemente $K(z) \ge 0$ para todo $z \in \mathbb{R}$ y puesto que θ' se anula para algunos valores, entonces el problema cambia su tipo de parabólico a elíptico. Para superar esto, empleamos un paso de regularización en (2.5). Dado $\epsilon > 0$ un pequeño parámetro de perturbación, reemplazamos la función no lineal θ en (2.5) por:

$$\theta_{\epsilon}(z) = \theta(z) + \epsilon z.$$

Se satisface que $\theta'_{\epsilon} > \epsilon$. Para obtener más detalles, consulte [32, 65]. De esta manera, las suposiciones sobre θ en (M1) y (M2) se aplican también a θ_{ϵ} . En este caso es posible obtener estimaciones de λ_{k+1}^n similares a las encontradas en este trabajo pero dependiendo del parámetro ϵ .

Observación 4.1.5. El L-scheme fue formulado originalmente para L constante [4, 15, 31, 43]. Consideramos que se pueden obtener mejores resultados numéricos cambiando la constante L en cada paso de tiempo, por lo que nosotros sugerimos que el parámetro L puede calcularse en cada paso de tiempo como:

$$L_{k+1} = \max\left\{\frac{1}{2}L_{\theta}, \frac{1}{\mu(\Omega)}\int_{\Omega} \theta'(\Psi_k)dx\right\}.$$

Es importante notar que si se usa L_k en lugar de L, todas afirmaciones hechas en el Teorema 4.1.1 siguen siendo válidas.

4.2 Esquema Type-Secant

Ahora formulamos un nuevo esquema inspirado en el método de la secante al que llamaremos **Type-Secant**. Este método es sólo localmente convergente, tiene un orden de convergencia superlineal y no requiere el cálculo de derivadas de las funciones involucradas. Según los experimentos numéricos presentados en esta tesis, este método ha demostrado ser más robusto que el método de *Newton* (ver ejemplo 2). El esquema es el siguiente: para $k \in \{0, 1, ..., N\}$ fijo,

$$\begin{cases} dados \ \Psi_{k}, \Psi_{k+1}^{n-1}, \Psi_{k+1}^{n} \in V_{h}, \ hallar \ \Psi_{k+1}^{n+1} \in V_{h} \ tal \ que : \\ \left\langle G_{k+1}^{n}(\Psi_{k+1}^{n}) \left(\Psi_{k+1}^{n+1} - \Psi_{k+1}^{n} \right), v \right\rangle + 2h_{t} \delta_{k+1}^{n} \left\langle K(\Psi_{k}) \nabla(\Psi_{k+1}^{n+1} + z), \nabla v \right\rangle \\ = 2\delta_{k+1}^{n} \left\langle \theta_{k} - \theta_{k+1}^{n} + h_{t} S_{k+1}, v \right\rangle \\ para \ todo \ v \in V_{h}, \end{cases}$$
(4.4)

donde G_{k+1}^n es una función real definida por $G_{k+1}^n(x) = \theta(x + \delta_{k+1}^n) - \theta(x - \delta_{k+1}^n)$ para todo $x \in \mathbb{R}$ y $\delta_{k+1}^n = \|\Psi_{k+1}^n - \Psi_{k+1}^{n-1}\|$. Comenzamos las iteraciones con $\Psi_{k+1}^0 = \Psi_k$ y Ψ_{k+1}^1 viene dada por la solución obtenida usando una iteración del *L*-scheme, por lo tanto $\|e_{k+1}^0\| < \|e_{k+1}^1\|$. A continuación analizaremos rigurosamente la convergencia del esquema propuesto. **Teorema 4.2.1.** Sea $k \in \{1, ..., N\}$ fijo. Si se satisfacen las hipótesis (M1), (M2), (M3) y Ψ_{k+1}^0 satisface la relación

$$\|\Psi_{k+1} - \Psi_{k+1}^0\| < \left[\frac{1}{\tilde{C}}\right]^{\frac{1}{\gamma}},$$
(4.5)

con

$$\tilde{C} = \frac{2C_p^2 C_\theta \sqrt{C\mu(\Omega)} \left[\mu(\Omega)^{-\frac{1}{2}\gamma} + 2(\gamma+1) \right]}{h_t K_r h^{\frac{d}{2}}(\gamma+1)},$$

entonces el esquema (4.4) converge en la norma de $H^1(\Omega)$ con un orden de convergencia de al menos $\frac{1}{2}(1+\sqrt{1+4\gamma})$.

Demostración. Restando (2.6) de (4.4), tomando $v = e_{k+1}^{n+1} := \Psi_{k+1}^{n+1} - \Psi_{k+1} \in V_h$, utilizando la hipótesis (M3), aplicando el teorema del valor medio en G_{k+1}^n y haciendo algunas operaciones algebraicas obtenemos

$$\left\langle \theta^{'}(a_{k+1}^{n})e_{k+1}^{n+1}, e_{k+1}^{n+1} \right\rangle + h_{t}K_{r} \|\nabla e_{k+1}^{n+1}\|^{2} \leq \left\langle \int_{\Psi_{k+1}^{n}}^{\Psi_{k+1}} \theta^{'}(t) - \theta^{'}(a_{k+1}^{n})dt, e_{k+1}^{n+1} \right\rangle$$

donde $a_{k+1}^n \in (\Psi_{k+1}^n - \delta_{k+1}^n, \Psi_{k+1}^n + \delta_{k+1}^n)$. Utilizando las desigualdades de Poincaré, Cauchy-Schwarz y Young, tenemos:

$$\frac{h_t K_r}{2C_p^2} \|e_{k+1}^{n+1}\|^2 + h_t K_r \|\nabla e_{k+1}^{n+1}\|^2 \le \frac{2C_p^2}{h_t K_r} \left\| \int_0^1 \left[\theta'(\alpha(z)) - \theta'(a_{k+1}^n) \right] \frac{d\alpha(z)}{dz} dz \right\|^2$$

donde $\alpha(z) = \Psi_{k+1}^n + z(\Psi_{k+1} - \Psi_{k+1}^n)$. Usando ahora la desigualdad de Cauchy-Schwarz, la estimación inversa dadas en el Lema 2.1.2 y la hipótesis (M2) obtenemos

$$\|e_{k+1}^{n+1}\|^2 + 2C_p^2 \|\nabla e_{k+1}^{n+1}\|^2 \le \frac{4C_p^4 C_\theta^2 C\mu(\Omega)^{1-\gamma}}{h_t^2 K_r^2 h^d} \left[\int_0^1 \|\alpha(z) - a_{k+1}^n\|^\gamma dz \right]^2 \|e_{k+1}^n\|^2.$$

Simplificando, la desigualdad anterior se convierte en

$$\begin{aligned} \|e_{k+1}^{n+1}\|^{2} + 2C_{p}^{2}\|\nabla e_{k+1}^{n+1}\|^{2} \\ \leq & \frac{4C_{p}^{4}C_{\theta}^{2}C\mu(\Omega)}{h_{t}^{2}K_{r}^{2}h^{d}} \left[\frac{\mu(\Omega)^{-\frac{1}{2}\gamma} + \gamma + 1}{\gamma + 1}\|e_{k+1}^{n}\|^{\gamma} + \|e_{k+1}^{n-1}\|^{\gamma}\right]^{2}\|e_{k+1}^{n}\|^{2}, \end{aligned}$$

$$(4.6)$$

de donde se consigue la siguiente desigualdad

$$\|e_{k+1}^{n+1}\| \leq \frac{2C_p^2 C_\theta \sqrt{C\mu(\Omega)}}{h_t K_r h^{\frac{d}{2}}} \left[\frac{\mu(\Omega)^{-\frac{1}{2}\gamma} + \gamma + 1}{\gamma + 1} \|e_{k+1}^n\|^{\gamma} + \|e_{k+1}^{n-1}\|^{\gamma}\right] \|e_{k+1}^n\|.$$

0

Tomemos en cuenta que de (4.5) se deduce la desigualdad $\tilde{C} ||e_{k+1}^0||^{\gamma} = \lambda < 1$ y puesto que en la primera iteración se usa el *L-scheme*, se tiene que $||e_{k+1}^1|| < ||e_{k+1}^0||$. Con estas consideraciones y usando el principio de inducción matemática se sigue inmediatamente que para todo n = 1, 2, ...

$$\|e_{k+1}^{n+1}\| \le \lambda \|e_{k+1}^n\|, \tag{4.7}$$

$$\|e_{k+1}^{n+1}\| \le \tilde{C} \|e_{k+1}^{n-1}\|^{\gamma} \|e_{k+1}^{n}\|.$$
(4.8)

De la desigualdad (4.7) se sigue la convergencia por lo menos lineal del esquema (4.4) en $L^2(\Omega)$ y usando la desigualdad (4.6) la convergencia del esquema en $H^1(\Omega)$ está demostrada. Ahora supóngase que existe una constante $r \in (1, 2)$ y K > 0 tal que

$$\|e_{k+1}^{n+1}\| \le \tilde{K} \|e_{k+1}^n\|^r, \tag{4.9}$$

de la desigualdad anterior se sigue que $||e_{k+1}^{n+1}|| \leq \tilde{K}^{r+1} ||e_{k+1}^{n-1}||^{r^2}$. Además usando (4.8) y (4.9) se obtiene $||e_{k+1}^{n+1}|| \leq \tilde{C}\tilde{K}||e_{k-1}^{n-1}||^{r+\gamma}$ y de las dos desigualdades anteriores es fácil deducir que $r = \frac{1}{2}(1 + \sqrt{1+4\gamma})$ y $\tilde{K} = \sqrt[r]{\tilde{C}}$. Esto demuestra que el orden de convergencia del esquema es por lo menos r.

Observación 4.2.1. Es fácil verificar que el método de Newton tiene un orden más alto que el esquema Type-Secant, puesto que $\frac{1}{2}(1 + \sqrt{1 + 4\gamma}) < 1 + \gamma$ para todo $\gamma > 0$. Sin embargo, los experimentos numéricos realizados muestran que el esquema Type-Secant es tan bueno como el método de Newton.

Observación 4.2.2. Como se mencionó anteriormente, el método Type-Secant tiene orden de convergencia superlineal, sin embargo solo convergente localmente. La desigualdad (4.5) nos muestra que la convergencia del esquema depende del tamaño de la malla espacial h, tal como sucede con el método de Newton, sin embargo el esquema parece ser más robusto que el método de Newton, como se muestra en el ejemplo 2 de este capítulo. Esto podría deberse a que la primera iteración del método es con el L-scheme, sin embargo el método Type-Secant también puede fallar.

4.3 Esquema L-Secant

El método *Type-Secant* tiene orden de convergencia superlineal, sin embargo, no es globalmente convergente por lo que el método podría fallar. Para aumentar la robustez del método *Type-Secant*, podemos usar algunas estrategias, por ejemplo, podemos utilizar un método mixto que consiste en realizar primero algunas iteraciones del método *L-scheme* y luego cambiar al *Type-Secant* [4]. Aún cuando este procedimiento debería proporcionar un esquema más robusto, es importante destacar que la convergencia de ninguno de estos métodos está garantizada. Es por esto que no profundizaremos en dicha metodología y alternativamente proponemos el mismo

procedimiento utilizado en la sección 4.1. Esto produce un esquema globalmente convergente con orden de convergencia superlineal.

En lo que sigue proponemos un nuevo método que lo llamamos L-Secant, este resulta de una combinación eficiente entre el L-scheme y el método Type-Secant, siguiendo la misma estrategia de la sección 4.1. Además demostramos la convergencia global del esquema. El esquema L-Secant es el siguiente: para $k \in \{0, 1, ..., N\}$ fijo,

Como en el apartado anterior $\lambda_{k+1}^n \in [0,1]$ es una sucesión en \mathbb{R} que converge a 1. Tenga en cuenta que si $\lambda_{k+1}^n = 1$ para todo n = 0, 1, ..., entonces el esquema se convierte en el *Type-Secant* y si por el contrario $\lambda_{k+1}^n = 0$ para todo n = 0, 1, ..., entonces es el método *L-scheme*.

Hipótesis 4.3.1. La siguiente hipótesis es necesaria para probar la convergencia:

 $\begin{array}{l} (M15) \ Para \ todo \ k \ fijo, \ la \ sucesión \ \left\{\lambda_{k+1}^n\right\}_{n\in\mathbb{N}} \ está \ definida \ como: \ \lambda_{k+1}^0 = 0 \ y \ para \\ n = 1, 2, \ldots \\ \\ \lambda_{k+1}^n = \left\{ \begin{array}{ll} 1, & si \quad B_{k+1}^n > 1, \\ \\ \zeta_{k+1}^n, & si \quad B_{k+1}^n \leq 1, \end{array} \right. \end{array}$

donde ζ_{k+1}^n es una sucesión tal que:

$$\zeta_{k+1}^n < B_{k+1}^n = \frac{(1+\gamma)h_t K_r h^{\frac{n}{2}}}{C_p^2 C_\theta \sqrt{C} \mu(\Omega)^{\frac{1+\gamma}{2}} \left[\|e_{k+1}^n\|^{\gamma} + (1+\gamma)\mu(\Omega)^{\frac{\gamma}{2}} \|\Psi_{k+1}^n - \Psi_{k+1}^{n-1}\|^{\gamma} \right]}$$

Teorema 4.3.1. Sea $k \in \{1, ..., N\}$ fijo. Supongamos que se satisfacen las hipótesis (M1), (M2), (M3) y (M15), entonces la sucesión de soluciones de (4.10) $\{\Psi_{k+1}^n\}_{n>0}$ converge a Ψ_{k+1} solución de (2.6) en el espacio $H^1(\Omega)$.

Demostración. La demostración combina las ideas utilizadas en la sección anterior con la técnica utilizada para la demostración del Teorema 4.1.1, por lo que no creemos necesario incluir la demostración. \Box

Observación 4.3.1. Ya que $||e_{k+1}^n||$ converge a cero, se tiene que luego de un número finito de iteraciones r, necesariamente $\lambda_{k+1}^n = 1$ para todo n > r Por lo tanto, deducimos que el esquema L-Secant tiene el mismo orden de convergencia que el esquema de Type-Secant. Por lo tanto, el L-Secant es un método globalmente convergente, con orden de convergencia superlineal y además no es necesario calcular las derivadas de las funciones involucradas.

4.4 Experimentos Numéricos

Mostraremos la eficiencia y robustez de los nuevos esquemas propuestos a través de tres ejemplo numéricos de referencia dados en la literatura y un ejemplo propuesto por nosotros. Presentamos resultados numéricos en dos y tres dimensiones. Para la discretización del tiempo, utilizamos el método de Euler semi-implícito. Para el espacio, elegimos elementos finitos lineales de Galerkin en mallas triangulares. Resumimos los resultados en varias tablas donde comparamos los resultados del *L*-scheme y el método de *Newton* con los nuevos esquemas de linealización. Todos los cálculos se realizaron en un portátil HP con un procesador Intel Core i5-7400 y 8 GB de RAM, y el software Freefem ++ v4.6.

4.4.1 Ejemplo 1

Este ejemplo fue usado en [15] para comparar el *L-scheme* con la solución exacta. Extendemos el ejemplo a tres dimensiones con una nueva solución analítica. Consideremos la siguiente ecuación diferencial parcial no lineal elíptica en $H^1(\Omega)$, $\Omega = (0,1)^3 \subset \mathbb{R}^3$:

$$\theta(\psi) - \Delta \Psi = f \quad \text{en} \qquad \Omega, \Psi = w \qquad \text{sobre} \quad \partial \Omega.$$
 (4.11)

La función no lineal $\theta \in C^1(\Omega)$ está definida por:

$$\theta(\Psi) = \begin{cases} \arctan(\Psi), & \text{si } \Psi < 1, \\ \frac{\pi}{4}, & \text{si } \Psi \ge 1. \end{cases}$$

El lado derecho f se elige de tal manera que la solución exacta de la ecuación diferencial (4.11) sea

$$\Psi(x, y, z) = w(x, y, z) = x^3 + y^3 - z^2 + x + y + \sin(\pi x)\sin(\pi y)\sin(\pi z).$$

El valor del parámetro L es el sugerido en [15], L = 1 y comenzamos las iteraciones con la solución inicial $\Psi_0(x, y, z) = 200$. Estudiamos dos tamaños de malla h = 0.2165y h = 0.0866. Comparamos las soluciones numéricas de los esquemas estudiados con la solución exacta. Para este efecto calculamos el error y el gradiente del error en la norma $L^2(\Omega)$ en cada iteración. Los resultados se muestran en la Tabla 4.1 y 4.2, con diez dígitos de precisión. En todos los esquemas notamos que los errores disminuyen rápidamente con cada iteración del método, para luego alcanzar un valor constante, esto se debe a que al inicio del proceso iterativo, el error de linealización es dominante. Pero luego el error de discretización es el dominante y la parte constante en las tablas presentadas significa que el error de discretización se ha alcanzado, es decir, el error de la linealización está por abajo de los 10 dígitos decimales considerados. Es importante resaltar que este fenómeno ya fue observado en [15]. Usamos los esquemas de linealización (4.1), (4.4) y (4.10) con $L = K_r = h_t = 1$. Puesto que disponemos de la solución exacta, calculamos en todos los casos $\|\Psi_n - \Psi\|$ y $\|\nabla\Psi_n - \nabla\Psi\|$ con 10 cifras decimales para las 10 primeras iteraciones. Los resultados de las simulaciones pueden encontrarse en la Tabla 4.1 y 4.2, donde lo primero que notamos es que todos los errores en la primera iteración de los tres esquemas propuestos son exactamente iguales al error del *L-scheme*. Esto no debería sorprendernos, pues la primera iteración de todos los esquemas propuestos es una iteración del *L-scheme*, sin embargo los errores en las restantes iteraciones son significativamente diferentes.

Los mejores resultados se obtienen con el esquema de Newton para las dos mallas investigadas, de hecho se alcanza el error de discretización en la cuarta iteración para la malla más gruesa y en la tercera iteración para los demás casos. Los resultados más pobres se obtienen con el *L-scheme*, donde el error de discretización para $||\Psi_n - \Psi||$ con la malla más gruesa se alcaza en la iteración 9, mientras que con la malla más fina se consigue en la iteración 10, lo mismo sucede con $||\nabla \Psi_n - \nabla \Psi||$ en la malla más gruesa, donde también se alcanza el error de discretización en la décima iteración. Sin embargo notamos que con la malla fina 10 iteraciones no son suficientes para alcanzar el error de discretización.

Los esquemas L-Newton y L-Secant son métodos de linealización robustos, puesto que su convergencia es independientemente de la solución inicial con que se comiencen las iteraciones, lo que significa gran diferencia con el método de Newton. Además estos métodos han demostrando un rendimiento muy superior al L-scheme. El L-Newton alcanza el error de discretización en la quinta iteración para las dos mallas y los dos errores investigados. El método Type-Secant es un esquema con orden de convergencia superlineal y que parece ser más robusto que el método de Newton, como podemos ver en el Ejemplo 2. Este esquema alcanza el error de discretización en iteración 5 para el caso de la malla más gruesa y en la iteración 4 para la malla más fina en los dos errores investigados. El esquema L-Secant alcanza el error de discretización en la sexta iteración en todos los casos, es decir, una iteración más que el esquema L-Newton.

4.4.2 Ejemplo 2

Este ejemplo fue propuesto en [4], es un problema en dos dimensiones espaciales y fue usado para estudiar comparativamente algunos esquemas de linealización. En esta tesis usamos este ejemplo para comparar los nuevos esquemas de linealización propuestos, fijamos una condición de parada y calculamos el número de iteraciones y el tiempo de cómputo [2, 4, 14, 15].

El ejemplo considera un dominio Ω en dos dimensiones espaciales (x,z) donde la coordenada z es hacia arriba. Ω esta dividido en dos partes: la zona no saturada $\Omega_v = (0, 1) \times (-3/4, 0)$ y la zona saturada $\Omega_s = (0, 1) \times (-1, -3/4)$. Las condiciones de contorno son mixtas, una condición de Dirichlet en $\Gamma_D = (0, 1) \times \{0\}$ y una condición

Iteraciones	Newton	L-scheme	L-Newton	Type-Secant	L-Secant
		h =	0.2165		
1	0.0007246223	4.7970496370	4.7970496370	4.7970496370	4.7970496370
2	0.0006949659	0.1542574307	0.1271016783	0.0007207487	0.1539303974
3	0.0006949658	0.0047842680	0.0006954915	0.0006949659	0.0043307553
4	0.0006949658	0.0006430956	0.0006949658	0.0006949658	0.0006949669
5	0.0006949658	0.0006927213	0.0006949658	0.0006949658	0.0006949658
6	0.0006949658	0.0006948921	0.0006949658	0.0006949658	0.0006949658
7	0.0006949658	0.0006949634	0.0006949658	0.0006949658	0.0006949658
8	0.0006949658	0.0006949658	0.0006949658	0.0006949658	0.0006949658
9	0.0006949658	0.0006949658	0.0006949658	0.0006949658	0.0006949658
10	0.0006949658	0.0006949658	0.0006949658	0.0006949658	0.0006949658
		h =	0.0866		
1	0.0001438592	4.7994184910	4.7994184910	4.7994184910	4.7994184910
2	0.0000418939	0.1546757905	0.1438479489	0.0001306337	0.1545448671
3	0.0000418939	0.0050353938	0.0000422515	0.0000418939	0.0048521581
4	0.0000418939	0.0001620467	0.0000418939	0.0000418939	0.0000418941
5	0.0000418939	0.0000412300	0.0000418939	0.0000418939	0.00004189 <mark>39</mark>
6	0.0000418939	0.0000418615	0.0000418939	0.0000418939	0.0000418939
7	0.0000418939	0.0000418928	0.0000418939	0.0000418939	0.0000418939
8	0.0000418939	0.00004189 <mark>38</mark>	0.0000418939	0.0000418939	0.0000418939
9	0.0000418939	0.000041893 <mark>9</mark>	0.0000418939	0.0000418939	0.0000418939
10	0.0000418939	0.0000418939	0.0000418939	0.0000418939	0.0000418939

Tabla 4.1: $\|\Psi_n-\Psi\|$ para el Ejemplo 1.

Iteraciones	Newton	L-scheme	L-Newton	Type-Secant	L-Secant
		<i>h</i> =	= 0.2165		
1	0.0087052492	27.3203555000	27.3203555000	27.3203555000	27.3203555000
2	0.0083799558	0.8425736642	0.6943374057	0.0086248652	0.8407891072
3	0.0083799558	0.0276544158	0.0083814541	0.0083799559	0.0252879608
4	0.0083799558	0.0083163229	0.0083799558	0.0083799558	0.0083799578
5	0.0083799558	0.0083763833	0.0083799558	0.0083799558	0.0083799558
6	0.0083799558	0.0083798377	0.0083799558	0.0083799558	0.0083799558
7	0.0083799558	0.0083799519	0.0083799558	0.0083799558	0.0083799558
8	0.0083799558	0.0083799557	0.0083799558	0.0083799558	0.0083799558
9	0.0083799558	0.008379955 <mark>8</mark>	0.0083799558	0.0083799558	0.0083799558
10	0.0083799558	0.0083799558	0.0083799558	0.0083799558	0.0083799558
		<i>h</i> =	= 0.0866		
1	0.0021973652	27.3399209000	27.3399209000	27.3399209000	27.3399209000
2	0.0005982302	0.8440751867	0.7849791253	0.0018729512	0.8433609081
3	0.0005982302	0.0274258843	0.0005997972	0.0005982302	0.0264288109
4	0.0005982302	0.0010534041	0.0005982302	0.0005982302	0.0005982307
5	0.0005982302	0.0005975775	0.0005982302	0.0005982302	0.0005982302
6	0.0005982302	0.0005981863	0.0005982302	0.0005982302	0.0005982302
7	0.0005982302	0.0005982287	0.0005982302	0.0005982302	0.0005982302
8	0.0005982302	0.0005982301	0.0005982302	0.0005982302	0.0005982302
9	0.0005982302	0.0005982301	0.0005982302	0.0005982302	0.0005982302
10	0.0005982302	0.0005982302	0.0005982302	0.0005982302	0.0005982302

Tabla 4.2: $\|\nabla \Psi_n - \nabla \Psi\|$ para el Ejemplo 1.

de Neumann sin flujo en el complemento $\Gamma_N = \partial \Omega - \Gamma_D$

$$\psi = -3$$
 sobre Γ_D y $[-K(\psi)\nabla(\psi + z)] \cdot \vec{n} = 0$ sobre Γ_N .

La condición inicial es discontinua en la transición de zona saturada a zona la vadosa y está dada por:

$$\psi(x, z, 0) = \psi_0(x, z) = \begin{cases} -3, & \text{si } (x, z) \in \Omega_v, \\ -z - \frac{3}{4}, & \text{si } (x, z) \in \Omega_s. \end{cases}$$

Seleccionamos un término fuente que toma valores positivos y negativos dado por:

$$f(x,z,t) = \begin{cases} -0.006\cos(\frac{3}{4}\pi z)\sin(2\pi x), & \text{si } (x,z) \in \Omega_v, \\ 0, & \text{si } (x,z) \in \Omega_s. \end{cases}$$

Las funciones hidráulicas están dadas mediante el modelo de Van-Genuchten, con los parámetros siguientes: $\alpha = 0.95$, n = 2.9, $\theta_s = 0.42$, $\theta_r = 0.026$, y $K_s = 0.12$. Note que la elección de n > 2 implica la Lipschitz continuidad de θ y K. Analizamos las soluciones numéricas después del primer paso de tiempo como es sugerido en [4]. Tomamos tres pasos de tiempo $h_t \in \{0.25, 1, 10\}$ con varios tamaños de malla h. Es fácil determinar que $L_{\theta} = 0.25$, usamos este valor en los métodos *L*-scheme y *L*2-scheme según lo establecido en la sección 3.3.2. Para detener las iteraciones, adoptamos el criterio $\|\psi_{k+1}^{n+1} - \psi_{k+1}^n\|_{L^2(\Omega)} \leq 10^{-10}$ siguiendo [2, 15]. Comenzamos las iteraciones con la solución obtenida en el paso tiempo anterior $\Psi_{k+1}^0 = \Psi_k$.

Los resultados se pueden ver en la Tabla 4.3. Lo primero que notamos es la falla del esquema de Newton, pues este no converge para las mallas más densas en absolutamente todos los pasos de tiempo investigados. Este resultado concuerda con la estimación dada en el Teorema 2.1.2, donde se afirman que la convergencia del método de Newton depende del tamaño de malla h. Todos los demás esquemas convergen para los tres pasos de tiempo y todos los tamaños de malla investigados. Observe que aún cuando el método Type-Secant es sólo localmente convergente, parece ser más robusto que el método de Newton, ya que converge para todos los pasos de tiempo y todas las mallas investigadas, aún en los casos donde el método de Newton falla. Esto podría deberse a que la primera iteración de este método corresponde al L-scheme. Aún así, recuerde que el Teorema 4.3.1 afirma que la convergencia del esquema también depende del tamaño de malla h por lo que si tomamos tamaños de malla más pequeños, el método podría fallar.

El número de iteraciones de cada uno de los métodos se puede observar en la Tabla 4.3 y en la Figura 4.1 únicamente para $h_t = 1$. El esquema de Newton cuando no falla y el método Type-Secant presentan números de iteración muy bajos, siendo el método Type-Secant un poco mejor, puesto que el esquema de Newton hace entre 6 y 10 iteraciones mientras el Type-Secant aproximadamente 6. El método L-Secant tiene un número de iteraciones que oscila entre 15 y 26, siendo ligeramente mayores

Número de iteresiones								<u>ae e</u> .			T	
		Nun	iero de	terac	iones				1 iemp)	
h	0.30	0.16	0.08	0.04	0.03	0.02	0.30	0.16	0.08	0.04	0.03	0.02
						$h_t = 0$.25					
Newton	6	7	7	10	falla	falla	0.11	0.55	2.1	11.77	falla	falla
L-scheme	145	133	124	123	123	123	1.84	2.12	26.9	106.43	167.19	417.49
L2-scheme	72	66	61	60	60	61	0.94	2.57	13.07	51.36	81.23	209.09
L-Newton	15	17	15	17	18	19	0.28	1.3	4.46	20.03	34.16	89.95
Type-Secant	6	6	6	6	6	7	0.10	0.43	1.70	6.71	10.51	31.93
L-Secant	20	21	20	21	22	23	0.35	1.60	6.05	25.33	41.35	110.03
						$h_t =$	1					
Newton	6	7	8	falla	falla	falla	0.13	0.55	2.51	falla	falla	falla
L-scheme	134	129	130	130	130	130	1.75	6.98	27.62	111.08	175.88	453.65
L2-scheme	66	65	64	64	64	64	0.86	3.52	13.66	53.77	86.8	218.44
L-Newton	10	14	14	15	16	17	0.19	1.06	4.18	17.59	30	81.78
Type-Secant	6	6	6	6	6	6	0.11	0.45	1.71	6.73	10.60	26.90
L-Secant	15	19	19	20	21	22	0.28	1.46	5.65	23.65	39.33	105.80
						$h_t = 1$	10					
Newton	6	7	falla	falla	falla	falla	0.12	0.53	falla	falla	falla	falla
L-scheme	96	113	117	117	118	118	1128.00	6.12	25.11	102.79	165.01	427.68
L2-scheme	51	61	60	60	60	61	0.67	3.30	12.65	50.44	82.00	212.10
L-Newton	8	9	10	11	12	13	0.15	0.73	3.01	13.01	22.26	62.87
Type-Secant	6	6	6	6	6	6	0.10	0.45	1.70	6.66	10.52	26.92
L-Secant	11	12	14	15	16	17	0.20	0.93	4.35	17.76	30.07	82.66

Tabla 4.3: Número de iteraciones y tiempo de CPU para el Ejemplo 2.



Figura 4.1: Número de iteraciones para el Ejemplo 2 con un paso de tiempo $h_t = 1$.



Figura 4.2: Tiempo de CPU para el Ejemplo 2 con un paso de tiempo $h_t = 1$.

que los registrados por el *L-Newton* que están entre 13 y 23, sin embargo tenga en cuenta que si el método de *Newton* converge lo hace más rápido. Notemos además que el número de iteraciones del *L-Newton* aumenta ligeramente en las mallas más densas. Los métodos *L-scheme* y *L2-scheme* producen los peores resultados en lo que respecta al número de iteraciones, siendo el *L-scheme* el peor de todos, en algunos casos hace hasta 8 veces más iteraciones que *L-Newton* y 20 más que el *Type-Secant*.

El tiempo de CPU se muestra en la Tabla 4.3 y la Figura 4.2 para $h_t = 1$, donde usamos 1/h en lugar de h en el eje horizontal con la finalidad de visualizar los resultados de mejor manera. Observamos que el tiempo de CPU más bajo lo registra el método de Newton cuando el esquema no falla, a pesar de que el esquema Type-Secant hace menos iteraciones. Esto se debe a que cada iteración del método Type-Secant es más



Figura 4.3: Evolucion de L_1^n para pasos de tiempo $h_t = 1$ con un tamaño de malla h = 0.02 Ejemplo 2.

costosa, puesto que hace más evaluaciones de la función θ , sin embargo, la diferencia en el tiempo de CPU entre ellos no es significativa. El esquema *L-Newton* obtiene un tiempo de cálculo menor que el *L-Secant* para todos los casos. El método *L-Newton* obtiene el menor tiempo de CPU entre todos los métodos que tienen convergencia global, mientras el esquema *L-Secant* posee el tiempo de CPU más pequeño entre todos los que tiene convergencia global y no utiliza la derivada de la función θ . El *L2-scheme* registra menos tiempo de CPU que el *L-scheme*, pero utiliza casi el doble de tiempo que el método *L-Newton*. En cualquiera de los casos, esto se debe a que la convergencia de los métodos *L-scheme* y *L2-scheme* es lenta y aún cuando según [4] las iteraciones no son muy caras, estos métodos no exhiben un buen rendimiento ya que tienen solo orden lineal y con tasa de convergencia muchas veces grandes.

Los buenos resultados que presenta el esquema *L-Newton* se deben a la aproximación de la función θ' por L_{k+1}^n en cada una de las iteraciones (ver Observación 4.1.1). Esta aproximación se da a través de la sucesión λ_{k+1}^n , que es una sucesión real monótona creciente, cuyo primer valor es cero y luego de un número finito de iteraciones toma el valor de 1 como se puede ver en la Figura 4.4. A manera de ejemplo, seleccionemos el paso de tiempo $h_t = 1$ y un tamaño de malla h = 0.02, en este caso el método de *Newton* falla, según se puede apreciar en la Tabla 4.3, sin embargo el método *L-Newton* es convergente. Los valores que toma λ_1^n en cada i-teración se encuentran en la Figura 4.4 y como se puede observar forman una sucesión creciente que comienza en 0 y termina en 1, de hecho en las primeras 10 iteraciones, el valor de λ_1^n es aproximadamente cero, por lo tanto el método es prácticamente el *L-scheme*,



Figura 4.4: Evolución de λ_1^n con $h_t = 1$ y h = 0.02, para el Ejemplo 2.

entre la iteración 9 y la 15. λ_1^n toma valores significativamente mayores que 0, pero sin alcanzar el valor de uno, por lo que el *L-Newton* es un esquema cuasi-Newton, donde θ' se aproxima por la función L_1^n . Esta aproximación se puede ver en la Figura 4.3, donde podemos notar que la función L_1^n en color azul se acerca a θ' en color rojo cuando λ_1^n se acerca a uno. A partir de la iteración 16, el esquema corresponde al método de *Newton*, puesto que $\lambda_1^n = 1$ para todo $n \ge 16$. En consecuencia tenemos un algoritmo robusto y eficiente y con un orden de convergencia $1 + \gamma$.

El esquema *L-Secant* funciona de forma parecida al método descrito anteriormente, es así como en la Figura 4.4, se observa la sucesión λ_1^n con los mismos parámetros usados para el *L-Newton*, al igual que el caso anterior λ_1^n es una sucesion monótona creciente que comienza en cero y luego de un número finito de iteraciones toma el valor de 1, de hecho para las primeras 13 iteraciones del *L-Secant* el método es prácticamente el *L-scheme* y a partir de la iteración 21 el método es el *Type-Secant*.

Para este ejemplo, el mejor método es el Type-Secant puesto que tiene mejor rendimiento que los otros esquemas y además no es necesario calcular las derivadas de las funciones involucradas. Aunque para algunos tamaños de malla, el esquema de Newton es mejor, este falla en varios casos, lo que hace que este método no sea recomendable. Debemos recordar que el esquema Type-Secant también es localmente convergente, lo que podría significar una desventaja, puesto que el método podría fallar. Es por ello que a pesar de que el L-Newton y el L-Secant tienen un rendimiento más bajo, nosotros recomendamos su uso, ya que la convergencia de estos esquemas está asegurada. El L-scheme y el L2-scheme tienen un rendimiento muy pobre, registran un alto número de iteraciones y un tiempo de CPU muy grande, en particular, el método L-scheme tiene una convergencia muy lenta para este ejemplo, registrando los peores tiempos de cálculo de entre todos los esquemas investigados.

4.4.3 Ejemplo 3

Este ejemplo ha sido usado por varios autores [3,22,26,27] y fue propuesto por primera vez en [3] para estudiar el método de linealización de *Picard Modificado*. Consideramos el caso de un suelo arenoso representado por el dominio espacial $\Omega = (0, 2) \times (-40, 0)$. Las funciones hidráulicas están dadas por un modelo Haverkamp [8]. Los parámetros del modelo vienen dados por: $\theta_s = 0.287$, $\theta_r = 0.075$, $\alpha = 1.611 \times 10^6$, $\beta = 3.96$, $K_s = 9.44 \times 10^{-3}$, $A = 1.175 \times 10^6$, y $\gamma = 4.74$. Al igual que el ejemplo anterior fijamos la condición de parada $\|\psi_{k+1}^{n+1} - \psi_{k+1}^n\|_{L^2(\Omega)} \leq 10^{-10}$ y calculamos el número de iteraciones y tiempo de CPU. No comenzamos las iteraciones de los métodos de linealización con la solución en el tiempo anterior como es habitual, sino que comenzamos todas las iteraciones con la función $\psi_{k+1}^0(x,z) = 1.02z - 20.7$. Analizamos las soluciones numéricas para los pasos de tiempo $h_t = 1, 10$ y los tamaños de malla h = 0.9428, 0.6248, 0.3195. Además usamos dos condiciones iniciales $\Psi_0 = -61.5$ y $\Psi_0 = -40$.

Los resultados de convergencia se muestran en la Tabla 4.4 para la primera condición inicial y en la Tabla 4.5 para la segunda condición inicial. Analizamos el número de iteraciones y el tiempo de CPU. Observamos que todos los esquemas convergen, inclusive el esquema de *Newton*, para todos las pasos de tiempo, tamaños de malla y las dos condiciones iniciales investigadas. Observamos que el tamaño de la malla no tiene un efecto significativo en el número de iteraciones que utilizan los métodos de linealización analizados.

El L2-scheme tiene mejor rendimiento que el L-scheme cuando usamos la condición inicial $\Psi_0 = -61.5$ en contraste con la segunda condición inicial donde tiene un mejor rendimiento el L-scheme. Esto se puede apreciar en la Tabla 4.4 y 4.5, si utilizamos $\Psi_0 = -61.5$ el L2-scheme usa 45 iteraciones, mientras el otro método tiene un desempeño muy pobre con 114 iteraciones. Por otro lado el L2-scheme con la segunda condición inicial usa 74 y 71 iteraciones respectivamente para los dos pasos de tiempo investigados, mientras el método L-scheme tiene un mejor desempeño obteniendo entre 17 y 19 iteraciones. Aún cuando las iteraciones de los métodos L-scheme y L2-scheme no son costosas, su rendimiento es muy pobre, esto se debe a que estos esquemas tienen una convergencia lenta caracterizada por una gran cantidad de iteraciones y tiempos de cálculo mayores que los esquemas propuestos en esta tesis.

El esquema Type-Secant muestra un gran rendimiento, registrando tiempos de CPU y número de iteraciones muy bajos, de hecho su rendimiento es muy similar al método de Newton, utilizando la primera condición inicial. El método Type-Secant hace 8 iteraciones y el método Newton 7 para todas las mallas utilizadas. Con la segunda condición inicial, el Type-Secant registra 7 iteraciones mientras que el método de Newton tiene 8 y 7 iteraciones para $h_t = 1$ y $h_t = 10$ respectivamente. El bajo número de iteraciones del esquema Type-Secant da como resultado tiempos de CPU más bajos y en algunos casos incluso más pequeños que los del método de Newton,

Número de iteraciones — Tiempo de CPU									
h		0 6248	0.3105		0 6248	0.310			
11	0.9420	0.0240	0.5195	0.9420	0.0240	0.3190			
$h_t = 1$									
Newton	7	7	7	0.69	2.10	7.82			
L-scheme	114	114	114	10.25	30.46	133.78			
L2-scheme	45	45	45	4.87	12.36	47.71			
L-Newton	21	22	23	2.13	6.45	25.66			
Type-Secant	9	8	8	0.84	2.17	8.24			
L-Secant	23	23	24	2.19	6.53	25.91			
		h_t	= 10						
Newton	7	7	7	0.70	2.03	7.80			
L-scheme	114	114	114	10.20	30.17	114.52			
L2-scheme	45	45	45	4.01	11.98	46.87			
L-Newton	17	18	19	1.69	5.24	21.31			
Type-Secant	8	8	8	0.73	2.18	8.20			
L-Secant	18	19	20	1.71	5.38	21.38			

Tabla 4.4: Iteraciones y tiempo de CPU para tamaños de malla de 0.9428, 0.6248, y 0.3195 usando un paso de tiempo de $h_t = 1, 10$ para la condición inicial $\Psi_0 = -61.5$ (Ejemplo 3)

como es el caso de la segunda condición inicial con $h_t = 1$.

Para este ejemplo el método Type-Secant es una extraordinaria alternativa al esquema de *Newton*, ya que tiene un desempeño similar, exhibiendo un número de iteraciones y tiempos de CPU muy similares al método de *Newton* y no usa la derivada de la función θ .

De todos los esquemas estudiados que poseen convergencia global, el L-Newton y L-Secant se convierten en una buena alternativa, estos esquemas registran prácticamente el mismo número de iteraciones y tienen el menor número de iteraciones de entre todos los esquemas globalmente convergentes analizados en este ejemplo.

El L-Newton y el L-Secant registran prácticamente el mismo número de iteraciones, de hecho, con la primera condición inicial registran entre 25 y 26 iteraciones para $h_t = 1$ y 23 o 24 iteraciones con $h_t = 10$, mientras que para la segunda condición inicial estos métodos tienen valores de 12 o 13 iteraciones para $h_t = 1$ y 11 para $h_t = 10$.

El tiempo de CPU se muestra en la Tabla 4.4 y 4.5 para las dos condiciones iniciales utilizadas, observamos que el tiempo de CPU más bajo lo registra el método de *Newton* en todos los casos. Puesto que el esquema *Type-Secant* hace un poco más de iteraciones que el método de *Newton*, su tiempo de cómputo es ligeramente más alto. El *L-scheme* tiene el tiempo de CPU más pobre para la primera condición inicial mientras que para $\Psi_0 = -40$ el tiempo de cómputo más grande lo registra el



Figura 4.5: Evolución de λ_{k+1}^n con $h_t = 1$, h = 0.3195 y $\Psi_0 = -40$, (Ejemplo 3).

Tabla 4.5: Iteraciones y tiempo de CPU para tamaños de malla de 0.9428, 0.6248, y 0.3195 usando un paso de tiempo de $h_t = 1, 10$ para la condición inicial $\Psi_0 = -40$ (Ejemplo 3).

	Númer	o de itera	aciones	Tiempo de CPU			
h	0.9428	0.6248	0.3195	0.9428	0.6248	0.3195	
		h_t	t = 1				
Newton	8	8	8	0.77	2.38	8.86	
L-scheme	17	18	19	1.70	5.28	20.66	
L2-scheme	74	74	74	6.66	21.21	78.90	
L-Newton	11	11	11	1.09	3.39	12.15	
Type-Secant	9	9	9	0.83	2.47	9.29	
L-Secant	12	12	13	1.15	3.39	14.03	
		h_t	= 10				
Newton	7	7	7	0.71	2.09	7.85	
L-scheme	19	19	19	1.84	5.45	20.56	
L2-scheme	71	71	71	7.86	19.98	79.58	
L-Newton	10	10	10	0.98	2.93	11.00	
Type-Secant	8	8	8	0.73	2.17	8.29	
L-Secant	11	11	12	1.06	3.12	12.91	



Figura 4.6: Geometría y condiciones de borde del Ejemplo 4.

L2-scheme. El método L-Newton obtiene prácticamente el mismo tiempo de cálculo que el L-Secant para todos los casos, siendo el del L-Newton ligeramente más bajo. Notemos que aún cuando los dos esquemas tienen prácticamente el mismo número de iteraciones, el tiempo de cálculo del L-Secant es ligeramente mayor que el del L-Newton, lo que significa que cada iteración es más cara. Además el L-Newton obtiene el menor tiempo de CPU entre todos los métodos que tienen convergencia global, mientras el esquema L-Secant posee el tiempo de CPU más pequeño entre todos los que tiene convergencia global y no utiliza la derivada de la función θ .

4.4.4 Ejemplo 4

Este ejemplo considera un tipo de suelo arenoso en el dominio $\Omega = (0, 2) \times (-2, 0) \subset \mathbb{R}^2$, que representa una sección vertical del subsuelo. Las funciones hidráulicas están dadas mediante un modelo Van-Genuchten con los parámetros: $\theta_s = 0.446, \theta_r = 0, \alpha = 0.152, n = 1.17, m = 1 - (1/n)$ y $K_s = 0.00082$. El problema usa condiciones de borde mixtas, la condición de Dirichlet está dada por $\Psi(x, z, t) = 2$ para todo $(x, z) \in \Gamma_{D1} = (0.75, 1.25) \times \{0\}$ y $\Psi(x, z, t) = -1.26$ para todo $(x, z) \in \Gamma_{D2} = (1, 2) \times \{-2\}$, en el resto de la frontera $\Gamma_N = \partial \Omega - (\Gamma_{D1} \cup \Gamma_{D2})$ se impone una condición de Neumann homogénea, ver Figura 4.6. La presión inicial es $\Psi(x, z, 0) = -1.26$ para todo $(x, z) \in \Omega$.

El criterio de parada para todos los métodos de linealizacion, está dado por $\|\psi_{k+1}^{n+1} - \psi_{k+1}^n\|_{L^2(\Omega)} \leq 10^{-10}$ y así mismo todos los esquemas investigados comienzan las iteraciones con la solución obtenida en el paso de tiempo anterior, es decir, $\psi_{k+1}^0 = \psi_k$. Examinamos las soluciones numéricas solamente para el primer paso de tiempo como fue sugerido en [4] para los pasos de tiempo $h_t = 0.25, 1, 5$ y los tamaños de malla h = 0.354, 0.155, 0.097.

Los resultados de convergencia se muestran en la Tabla 4.6, lo primero que notamos

	Núme	ro de ite	eraciones	Tien	npo de (CPU			
h	0.354	0.155	0.097	0.354	0.155	0.097			
$h_t = 0.25$									
Newton	4	5	5	0.21	1.58	3.93			
L-scheme	14	13	14	0.55	2.86	7.96			
L2-scheme	112	121	122	4.16	26.40	70.49			
L-Newton	7	7	7	0.39	2.20	5.64			
Type-Secant	5	5	5	0.25	1.43	3.72			
L-Secant	8	8	8	0.43	2.49	6.40			
$h_t = 1$									
Newton	5	5	5	0.26	1.52	3.95			
L-scheme	13	16	15	0.50	3.49	8.50			
L2-scheme	119	122	112	4.48	26.63	69.64			
L-Newton	6	7	7	0.32	2.20	5.64			
Type-Secant	5	5	5	0.25	1.44	3.71			
L-Secant	7	8	8	0.39	2.45	6.43			
		h_t	= 5						
Newton	5	5	5	0.13	1.57	4.23			
L-scheme	13	14	14	0.50	3.06	7.97			
L2-scheme	93	96	97	3.44	21.12	55.48			
L-Newton	5	6	6	0.28	1.89	4.84			
Type-Secant	6	5	5	0.29	1.46	3.71			
L-Secant	7	7	7	0.40	2.18	5.62			

Tabla 4.6: Número de iteraciones y tiempo de CPU para tamaños de malla h de 0.354, 0.155 y 0.097 y pasos de tiempo h_t de 0.25, 1 y 5. (Ejemplo 4).

es que todos los esquemas convergen, incluido el esquema de *Newton*, para todas las mallas y pasos de tiempo investigados. Sin embargo, el esquema de *Newton* podría fallar para tamaños de malla más pequeños.

En la Tabla 4.6 se muestra el número de iteraciones para los esquemas estudiados. Los mejores resultados los obtienen los esquemas de Newton y Type-Secant con 5 iteraciones aproximadamente, sin embargo, el método L-Newton tiene un rendimiento similar con 6 iteraciones aproximadamente, seguido del L-Secant con 7 iteraciones. En cuanto al L-scheme, hace entre 13 y 16 iteraciones, es decir, aproximadamente el doble de los esquemas L-Newton y L-Secant. El método L2-scheme tiene un rendimiento muy pobre, es demasiado lento, registrando 120 iteraciones aproximadamente, aunque tiene una disminución significativa cuando $h_t = 5$ con aproximadamente 95 iteraciones.



Figura 4.7: Perfiles de presión para T = 20, $h_t = 1$ y un tamaño de malla h = 0.155 (Ejemplo 4).

En la Tabla 4.6 se muestran también los tiempos de CPU, podemos observar que el esquema más rápido es el *Newton*, sin embargo, recordemos que este esquema puede fallar, el esquema *Type-Secant* muestra tiempos de cómputo muy parecidos al *Newton*, a veces incluso más pequeños, dado que este método parece ser más robusto que el esquema de *Newton* se convierte en una buena alternativa. Por otro lado el método *L-Newton* muestra tiempos de CPU un poco más altos que el método de *Newton* y muy similares al *L-Secant*, observe que el método *L-Secant* registra tiempos de cómputo ligeramente más altos que el *L-Newton*. En cuanto al *L-scheme* este muestra tiempos de CPU aproximadamente 30% más grandes que el *L-Newton*. Como esperábamos, el *L2-scheme*, muestra un tiempo de cálculo muy pobre en comparación con todos los otros esquemas.

El esquema L-Newton es el mejor método para este ejemplo, ya que su rendimiento es apenas superado por el método de Newton. Creemos que es un pequeño precio a cambio de un algoritmo mucho más robusto, cuya convergencia está garantizada para cualquier solución inicial seleccionada. En el caso de los esquemas que no utilizan la derivada de θ , el esquema Type-Secant es sin duda el mejor. Este esquema hace prácticamente el mismo número de iteraciones que el método de Newton con tiempos de cómputo similares. Sin embargo, recuerde que el esquema Type-Secant solo tiene convergencia local, por lo que el método podría fallar, aún cuando parece ser más robusto que el esquema de Newton. Es por esta razón que preferimos el método L-Secant en lugar del Type-Secant aún cuando tiene un rendimiento ligeramente menor. En este ejemplo, aún cuando el L-scheme hace muchas iteraciones, aproximadamente 14, su tiempo de cálculo es aceptable puesto que sus iteraciones no son costosas. El peor método para este ejemplo es sin lugar a dudas, el L2-scheme, puesto que tiene un rendimiento muy pobre y note que incluso cuando sus iteraciones no son costosas, su tiempo de CPU es muy grande.

4.5 Conclusión

En este capítulo consideramos algunos métodos de linealización para la Ecuación de Richards. Los métodos se estudiaron comparativamente, para lo cual examinamos el número de iteraciones y el tiempo CPU. Se discuten los esquemas iterativos para resolver los sistemas no lineales totalmente discretos obtenidos luego de aplicar un Esquema de Euler y el método elementos finitos de Galerkin a la ecuación diferencial parcial no lineal degenerada.

Nos concentramos en el método de Newton y el L-scheme y proponemos tres nuevos esquemas de linealización, el L-Newton, Type-Secant y L-Secant. Proporcionamos una prueba rigurosa de la convergencia de estos métodos. Todos los esquemas propuestos tienen un orden de convergencia superlineal, algunos de estos esquemas pueden tener incluso convergencia cuadrática como es el caso del L-Newton. Dos de estos métodos tienen convergencia global y el restante parece ser más robusto que el esquema de Newton. Presentamos cuatro ejemplos numéricos ilustrativos, con parámetros realistas, en 2 y 3 dimensiones espaciales

Los hallazgos numéricos apoyan los resultados teóricos encontrados. Los experimentos numéricos muestran que el esquema de Newton cuando no falla, posee un buen rendimiento ya que requiere un menor número de iteraciones y tiempo de CPU, sin embargo como ya dijimos es el menos robusto de todos los esquemas investigados, ya que existen casos en los que el método falló. Aún cuando el L-scheme es más fácil de implementar, no necesita el cálculo de la derivada de θ y las iteraciones son más económicas, este solo tiene una convergencia lineal, lo que podría resultar en una convergencia muy lenta. Además no existe en la literatura una técnica adecuada para seleccionar el parámetro L y los ejemplos numéricos presentados en este capítulo muestran que diferentes selecciones del parámetro L pueden conducir a esquemas con rendimientos diferentes. Por otro lado, los esquemas L-Newton y L-Secant son muy robustos, por lo que su convergencia está garantizada, su rendimiento es ligeramente inferior al método Newton y muy superior al L-scheme y al L2-scheme. Es por esta razón que estos dos métodos son una gran alternativa al método de Newton. En los cuatro ejemplos discutidos, el esquema L-Secant mostró un desempeño similar al método de *Newton* en lo que se refiere al número de iteraciones y tiempos de CPU, además este esquema también demostró ser más robusto que el método de Newton ya que converge donde el esquema de Newton falló. Los esquemas propuestos en este capítulo constituyen una alternativa eficaz a los tradicionales.

Capítulo 5 CONCLUSIONES Y RECOMENDACIONES

Esta tesis propone ocho nuevos esquemas de linealización para la Ecuación de Richards, se prueba teóricamente sus propiedades de convergencia y se presentan varios ejemplos numéricos en dos y tres dimensiones que confirman los hallazgos teóricos. Los ejemplos nos permiten estudiar el rendimiento de los métodos. Los nuevos esquemas son comparados con el método de *Newton*, que se caracteriza por tener un orden de convergencia superlineal pudiendo incluso ser cuadráticamente convergente, pero solo es localmente convergente, por lo que el esquema puede fallar. Más aún, la convergencia del método depende del tamaño de la malla espacial como se muestra en la sección 2.1 de esta tesis. Los nuevos métodos también son comparados con el *L-scheme*, caracterizado por ser globalmente convergente. Sin embargo, el método es muy lento con un rendimiento muy pobre.

Para resumir, este trabajo se concentra en los métodos linealización para la Ecuación de Richards y las contribuciones de esta tesis son:

- Basado en las ideas de [4,15,32,41,43] se prueba formalmente algunas propiedades de convergencia de los métodos de *Newton*, *Picard Modificado* y *L-scheme* para la Ecuación de Richards con una discretización de elementos finitos de Galerkin.
- Se propone una generalización del *L-scheme* y se demuestra su convergencia global bajo ciertas hipótesis de la función \widehat{L}_{k+1}^n .
- Se proponen cuatro nuevos esquemas numéricos globalmente convergente y con orden lineal, dos de estos métodos no usan la derivada de la función θ .
- Se proponen cuatro nuevos esquemas de linealización con orden superlineal, en dos de estos se demuestra su convergencia global y se obtiene el orden de convergencia teórico, en lo dos restantes solo damos el orden de convergencia teórico, dos de estos métodos no usan la derivada de la función θ .

5.1 Conclusiones

- Si usamos la Ecuación de Richards en su cuarta forma con un Esquema de Euler totalmente implícito en el tiempo y el método de elementos finitos de Galerkin en el espacio, los métodos de Newton y Picard Modificado convergen solo localmente, su convergencia local solo puede ser demostrada si previamente se usa un paso de regularización sobre la función θ . Además el método de Picard Modificado teóricamente es más robusto que el método de Newton, que posee un orden de convergencia superlineal, incluso cuadrática en algunos casos, mientras la convergencia del Picard Modificado es solo Lineal. En los dos métodos la convergencia depende del tamaño de la malla espacial.
- Si usamos la Ecuación de Richards en su forma mixta con un Esquema de Euler semi-implicito en el tiempo y elementos finitos de Galerkin en el espacio, el método de Newton también tiene convergencia únicamente local. Sin embargo para probar la convergencia del esquema no es necesaria la regularización de θ. La condición de convergencia en este caso sigue dependiendo del tamaño de la malla espacial y en este caso también depende del paso de tiempo, la convergencia sigue siendo superlineal.
- Si utilizamos la forma mixta y cuarta forma con los esquemas de Euler totalmente implícito y semi-implícito en el tiempo y elementos finitos de Galerkin en el espacio, en todos los casos se puede demostrar la convergencia global del *L-scheme*, con la única diferencia que cuando se utiliza el Esquema de Euler totalmente implícito es impresindible una restricción en el paso de tiempo, además no es necesario en ningún caso usar una regularización sobre la función θ . Este esquema tiene un orden de convergencia solamente lineal.
- Los nuevos esquemas de linealización presentados en esta tesis son más robustos que el método de *Newton*, mucho más eficientes que el *L-scheme* y se convierten en una valiosa alternativa cuando se trata de resolver la Ecuación de Richards.
- En el capítulo tres proponemos una generalización del *L*-scheme y demostramos su convergencia global bajo ciertas hipótesis de la función \widehat{L}_{k+1}^n . Esta nueva teoría desarrollada permite formular nuevos esquemas numéricos con convergencia global, propusimos en esta tesis cinco nuevos esquemas. Es evidente que se puede utilizar estos hallazgos teóricos para desarrollar otros esquemas de linealización.
- Los experimentos numéricos presentados en el capítulo tres confirman los hallazgos teóricos del capítulo y además sugieren que cuatro de los esquemas que poseen orden de convergencia lineal tienen un mejor rendimiento que el *L*scheme, ya que registran un menor número de iteraciones, tiempos de CPU más bajos y tasas de convergencia más pequeñas. El esquema numérico restante es

muy rápido y tiene el mismo orden de convergencia que el método de *Newton* y aún cuando no se logró demostrar totalmente la convergencia del método, los experimentos numéricos muestran que es más robusto que el método de *Newton* e inclusive en ocasiones puede tener un mejor rendimiento.

- En el capítulo 4 presentamos tres nuevos esquemas con orden de convergencia superlineal, probamos formalmente la convergencia global y el orden de convergencia de dos de ellos. Para el restante esquema demostramos formalmente su convergencia local y hallamos su orden de convergencia. Los experimentos numéricos realizados en este capítulo ratifican estos hallazgos teóricos.
- El esquema *L-Newton* es un esquema globalmente convergente y tiene el mismo orden de convergencia del método de *Newton*. Aún cuando en la práctica es un poco más lento que el esquema *Newton*, creemos que es un pequeño precio a cambio de la convergencia global del método.
- El Type-Secant es un esquema localmente convergente, los experimentos numéricos muestran que es mucho más robusto que el método de Newton y no usa la derivada de la función θ . El Type-Secant tiene un orden de convergencia superlineal, teóricamente menor que el método de Newton, sin embargo en la práctica tiene un rendimiento muy similar.
- El L-Secant es un método globalmente convergente que posee el mismo orden de convergencia teórico que el método Type-Secant. Aún cuando los experimentos numéricos muestran que es más lento que el método Type-Secant, pensamos que es un pequeño precio que se debe pagar a cambio de la convergencia global. El L-Secant es un método que no usa la derivada de la función θ, lo cual significa una gran ventaja del método.

5.2 Recomendaciones

• Dado que la convergencia del método de Newton depende del tamaño de la malla espacial y del paso de tiempo, (ver seccion 2.1.2), pensamos que se puede usar la estimación (2.8) para desarrollar un método multimalla con paso de tiempo adaptativo, para aumentar la robustez y el desempeño del método de Newton. Así, podríamos comenzar las iteraciones con una malla gruesa y luego de algunas iteraciones cambiar a una malla más fina. Esta técnica combinada con el uso de un paso de tiempo adaptativo podría llevar a la convergencia del método de Newton. Pensamos que al usar este procedimiento, no solamente el método de Newton será más robusto, sino que sería mucho más rápido, pues la mayor parte de las iteraciones deberían realizarse con la malla gruesa, lo que significa que los sistemas de ecuaciones serán mucho más pequeños y por tanto el método sera mucho mas rápido.

• Recomendamos estudiar estas mismas técnicas para otro tipo de discretizaciones espaciales, como el método de volúmenes finitos y así mismo para otro tipo de problemas más generales como los flujos en medios porosos multifásicos.

Bibliografía

- [1] Chen X. and Y. Dai. An approximate analytical solution of Richards' equation with finite boundary. *Boundary Value Problems*, 167:11p, 2017.
- [2] C. Paniconi and M Putti. A comparison of Picard and Newton iteration in the numerical solution of multidimensional variably saturated flow problems. *Water Resources Research*, 30:3357–3374, 1994.
- [3] M. Celia and E Bouloutas. General mass-conservative numerical solutions for the unsaturated flow equation. *Water Resources Research*, 26:1483–1496, 1990.
- [4] F. List and F. Radu. A study on iterative methods for solving Richards' equation. Computational Geosciences, 20:341–35, 2016.
- [5] G. Manoli, S. Bonetti, J. Domec, M. Putti, and G. Katul. Tree root systems competing for soil moisture in a 3D soil–plant model. Advances in Water Resources, 66:32–42, 2014.
- [6] G. Marinoschi. Functional approach to nonlinear models of water flow in soils. Mathematical Modelling: Theory and Applications. Springer, 1st edition. edition, 2010.
- M.Th. van Genuchten. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal, 44:892– 898, 1980.
- [8] R. Haverkamp, M. Vauclin, J. Touma, P. J. Wierenga, and G. Vachaud. A comparison of numerical simulation models for one-dimensional infiltration. *Soil Science Society of America Journal*, 41:285–294, 1977.
- [9] J.W. Delleur. *The handbook of groundwater engineering*. CRC Press, second edition edition, 2006.
- [10] E. Buckingham. Studies on the movement of soil moisture. U.S. Department of Agriculture Bureau of Soils, 38, 1907.
- [11] L.A. Richards. Capillary conduction of liquids through porous mediums. *Journal of Applied Physics*, 1:318–333, 1931.

- [12] H.W. Alt and S. Luckhaus. Quasilinear elliptic-parabolic differential equations. Mathematische Zeitschrift, 183:311–341, 1983.
- [13] O. Misiats and K. Lipnikov. Second-order accurate monotone finite volume scheme for Richards' equation. *Journal of Computational Physics*, 239:123–137, 2013.
- [14] F. Lehmann and P. Ackerer. Comparison of iterative methods for improved solutions of the fluid flow equation in partially saturated porous media. *Transport* in Porous Media, 31:275–292, 1998.
- [15] M. Slodicka. A robust and efficient linearization scheme for doubly non-linear and degenerate parabolic problems arising in flow in porous media. *Journal on Scientific Computing*, 23:1593–1614, 2002.
- [16] M. Kuraz, P. Mayer, and T. Dagmar. An adaptive time discretization of the classical and the dual porosity model of Richards' equation. *Journal of Computational and Applied Mathematics*, 233:3167–3177, 2010.
- [17] B. Cumming, T Moroney, and I. Turner. A mass conservative control volume finite element method for solving Richards' equation in heterogeneous porous media. *Numerische Mathematik*, 51:845–864, 2011.
- [18] C. Zambra, M. Dumbser, E. Toro, and N. Moraga. A novel numerical method of high-order accuracy for flow in unsaturated porous media. *International Journal* for Numerical Methods in Engineering, 89:227–240, 2012.
- [19] Caviedes D., Garcia P., and J. Murillo. Conservation, stability and efficiency of a finite volume method for the 1D Richards' equation. *Journal of Hydrology*, 480:69–84, 2013.
- [20] H. Cao, T. Yu, and X. Yue. A Completely Discrete Heterogeneous Multiscale Finite Element Method for Multiscale Richards' Equation of Van-Genuchten-Mualem Model. *Journal of Applied Mathematics*, 2014:1–10, 2014.
- [21] W. Lai and F.L. Ogden. A mass-conservative finite volume predictor-corrector solution of the 1D Richards' equation. *Journal of Hydrology*, 523:199–127, 2015.
- [22] Y. Baron, V. Coudiére and P. Sochala. Adaptive multistep time discretization and linearization based on a posteriori error estimates for the Richards' equation. *Applied Numerical Mathematics. An IMACS Journal*, 112:104–125, 2017.
- [23] M. S. Slam. Sensitivity analysis of unsaturated infiltration flow using head based finite element solution of Richards' equation. *Matematika*, 3:131–148, 2017.

- [24] B. Deng and J. Wang. Saturated-unsaturated groundwater modeling using 3D Richards' equation with a coordinate transform of nonorthogonal grids. *Applied Mathematical Modelling*, 50:39–52, 2017.
- [25] R.S. Carsel and R.S. Parrish. Developing joint probability distributions of soil water retention characteristics. *Water Resources Research*, 24:755–769, 1988.
- [26] R. Eymard, M. Gutnic, and D. Hilhorst. The finite volume method for Richards' equation. Computational Geosciences, 3:259–294, 1999.
- [27] K. Brenner, D. Hilhorst, and H. Vu-Do. A gradient scheme for the discretization of Richards' equation. *Finite Volumes for Complex Applications VII-Elliptic, Parabolic and Hyperbolic Problems*, 1:537–545, 2016.
- [28] K. Brenner, D. Hilhorst, and H. Vu-Do. The generalized finite volume SUSHI scheme for the discretization of Richards' equation. *Vietnam Journal of Mathematics*, 44:557–586, 2017.
- [29] M. Berardia and M. Vurroa. The numerical solution of Richards' equation by means of method of lines and ensemble Kalman filter. *Mathematics and Computers in Simulation*, 125:38–47, 2016.
- [30] L. Fengnan, Y. Fukumoto, X. Zhao, I.S. Pop, and F. Radu. A Linearized Finite Difference Scheme for the Richards Equation Under Variable-Flux Boundary Conditions. *Journal of Scientific Computing*, 83:16, 2020.
- [31] F. Radu, I. Pop, and P. Knabner. Order of convergence estimates for an Euler implicit, mixed finite element discretization of Richards' equation. *SIAM Journal* on Numerical Analysis, 42:1452–1478, 2004.
- [32] I. Pop and F. Radu. Mixed Finite elements for the Richards' equation: linearization procedure. Journal of Computational and Applied Mathematics, 168:365– 373, 2004.
- [33] F.A. Radu, I.S. Pop, and P. Knabner. Error estimates for a mixed finite element discretization of some degenerate parabolic equations. *Numerische Mathematik*, 109:285–311, 2008.
- [34] C.J. Van Duyn and L.A. Peletier. Nonstationary filtration in partially saturated porous media. Archive for Rational Mechanics and Analysis, 78:173–198, 1982.
- [35] J. Kačur. On a solution of degenerate elliptic-parabolic systems in Orlicz-Sobolev spaces I. Mathematische Zeitschrift, 203:153–171, 1990.
- [36] F. Otto. L₁-contraction and uniqueness for quasilinear elliptic-parabolic equations. Journal of Differential Equations, 131:20–38, 1996.
- [37] T. Arbogast, M.M. Obeyesekere, and M.F. Wheeler. Numerical methods for the simulation of flow in root-soil systems. Society for Industrial and Applied Mathematics, 30:1677–1702, 1993.
- [38] C. Fassino and G. Manzini. Fast-secant algorithms for the non-linear Richards' equation. Communications in Numerical Methods in Engineering, 14:921–930, 1998.
- [39] L. Bergamaschi and M. Putti. Mixed finite elements and Newton type linearizations for the solution of Richards' equation. *International Journal for Numerical Methods in Engineering*, 45:1025–1046, 1999.
- [40] I. Pop. Error estimates for a time discretization method for the Richards' equation. Computational Geosciences, 6:141–160, 2002.
- [41] F.A. Radu, I.S. Pop, and P. Knabner. On the convergence of the Newton method for the mixed finite element discretization of a class of degenerate parabolic equation. *Numerical Mathematics and Advanced Applications*, 42:1192–1200, 2006.
- [42] H. Berninger, Kornhuber R., and O. Sander. Fast and robust numerical solution of the Richards' equation in homogeneous soil. SIAM Journal on Numerical Analysis, 49:2576–2597, 2014.
- [43] J.W. Both, K. Kumar, J.M. Nordbotten, I.S. Pop, and F. Radu. Linear iterative schemes for doubly degenerate parabolic equations. Advances in Water Resources, 94:11–22, 2018.
- [44] D. Michael, C.T. Tocci, and Kelley. Accurate and economical solution of the pressure-head form of Richards' equation by the method of lines. Advances in Water Resources, 20:1–14, 1997.
- [45] G. Williams and C. Miller. An evaluation of temporally adaptive transformation approaches for solving Richards' equation. Advances in Water Resources, 22:831– 840, 1999.
- [46] D. Kavetski, P. Binning, and S. Sloan. Adaptive time stepping and error control in a mass conservative numerical solution of the mixed form of Richards' equation. Advances in Water Resources, 24:595–605, 2001.
- [47] M. Kuraz, P. Mayer, and P. Pech. Solving the nonlinear and nonstationary Richards' equation with two-level adaptive domain decomposition (*dd*adaptivity). Applied Mathematics and Computation, 267:207–222, 2015.
- [48] K. Brenner and C. Cancés. Improving Newton's method performance by parametrization: the case of the Richards' equation. SIAM Journal on Numerical Analysis, 55:1760–1785, 2017.

- [49] D. Illiano, S. Pop, and S. Radu. Iterative schemes for surfactant transport in porous media. *Computational Geosciences*, page 18, 2020.
- [50] R. Timsina, H. Khanal, and K. Uprety. An Explicit Stabilized Runge–Kutta–Legendre Super Time-Stepping Scheme for the Richards Equation. *Mathematical Problems in Engineering*, 2021:1–11, 2021.
- [51] P. G. Ciarlet. The finite element method for elliptic problems, volume 2 of Classics in Applied Mathematics. SIAM, second edition edition, 1978.
- [52] S. C. Brenner and L. R. Scott. The mathematical theory of finite element methods, volume 15 of Texts in Applied Mathematics. Springer-Verlag, third edition edition, 1991.
- [53] J. Richard, S. Cheung, and T. Mai. Multiscale simulations for multi-continuum Richards equations. *Journal of Computational and Applied Mathematics*, 397:28, 2021.
- [54] E. Schneid, P. Knabner, and F. Radu. A priori error estimates for a mixed finite element discretization of the Richards' equation. *Numerische Mathematik*, 98:353–370, 2004.
- [55] I. S. Pop, M. Sepúlveda, F. A. Radu, and O. P. Vera Villagrán. Error estimates for the finite volume discretization for the porous medium equation. *Journal of Computational and Applied Mathematics*, 234:2135–2142, 2010.
- [56] P. Knabner and L. Angermann. Numerical methods for elliptic and parabolic partial differential equations. Springer, 2003.
- [57] M. Grau. On iterative methodsto solve nonlinear equations. PhD thesis, Universitat Politècnica de Catalunya, España, 2015.
- [58] J. E. Dennis and R. B. Schnabel. Numerical Methods for Unconstrained Optimization and Nonlinear Equations, volume 16 of Classics in Applied Mathematics. SIAM, 1st edition, 1996.
- [59] L.C. Evans. *Partial differential equations*. American Mathematical Society, second edition edition, 2010.
- [60] Lipnikova K., D. Moulton, and D. Svyatskiy. New preconditioning strategy for Jacobian-free solvers for variably saturated flows with. Advances in Water Resources, 94:11–22, 2016.
- [61] R. Younis, A. Tchelepi, T. Dagmar, and K. Aziz. Adaptively localized continuation-Newton method nonlinear solvers that converge all the time. SPE Journal, 15:526–544, 2010.

- [62] X. Wang and H. Tchelepi. Trust-region based solver for nonlinear transport in heterogeneous porous media. J. Comput. Phys., 253:114–137, 2013.
- [63] X. C. Cai and D. E. Keyes. Nonlinearly preconditioned inexact Newton algorithms. *SIAM Journal on Scientific Computing*, 24:183–200, 2002.
- [64] V. Dolean, M. J. Gander, W. Kheriji, F. Kwok, and R. Masson. Nonlinear preconditioning: how to use a nonlinear Schwarz method to precondition Newton's Method. SIAM J. Scientific Computing, 6:3357–3380, 2016.
- [65] I. S. Pop and B. Schweizer. Regularization schemes for degenerate Richards' equations and outflow conditions. *Mathematical Models and Methods in Applied Sciences*, 21:1685–1712, 2011.
- [66] H. Brézis. Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. Mathematics studies. North-Holland; American Elsevier, 1st edition edition, 1973.