

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



ĐỖ HẢI SƠN

**NGHIÊN CỨU NHẬN DẠNG HỆ THỐNG
VỚI TRI THỨC MỚI CHO HỆ THỐNG TRUYỀN THÔNG
MIMO KÍCH THƯỚC LỚN**

**LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ - VIỄN THÔNG**

Hà Nội - 2023

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

ĐỒ HẢI SƠN

NGHIÊN CỨU NHẬN DẠNG HỆ THỐNG
VỚI TRI THỨC MỚI CHO HỆ THỐNG TRUYỀN THÔNG
MIMO KÍCH THƯỚC LỚN

Ngành: Công nghệ Kỹ thuật Điện tử - Viễn thông

Chuyên ngành: Kỹ thuật Viễn thông

Mã số: 8510302

LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ - VIỄN THÔNG

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. Trần Thị Thúy Quỳnh

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

ĐỖ HẢI SƠN

NGHIÊN CỨU NHẬN DẠNG HỆ THỐNG
VỚI TRI THỨC MỚI CHO HỆ THỐNG TRUYỀN THÔNG
MIMO KÍCH THƯỚC LỚN

Ngành: Công nghệ Kỹ thuật Điện tử - Viễn thông

Chuyên ngành: Kỹ thuật Viễn thông

Mã số: 8510302

TÓM TẮT LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ - VIỄN THÔNG

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn thạc sĩ **Nghiên cứu nhận dạng hệ thống với tri thức mới cho hệ thống truyền thông MIMO kích thước lớn** là công trình nghiên cứu thực sự của tôi, được thực hiện dựa trên cơ sở lý thuyết, kiến thức chuyên ngành dưới sự hướng dẫn khoa học của TS. Trần Thị Thúy Quỳnh.

Tôi xin cam đoan những công việc trong luận văn thực hiện chưa từng được các tác giả khác đề xuất. Với sự hiểu biết của mình, tôi chắc chắn các số liệu, kết quả trong luận văn là trung thực và chưa được công bố ở đâu và trong bất cứ công trình nào trừ công trình của tác giả và tài liệu tham khảo.

Nếu có gì sai trái, tôi xin hoàn toàn chịu trách nhiệm.

Hà Nội, ngày ... tháng ... năm 2023

Học viên

Đỗ Hải Sơn

LỜI CẢM ƠN

Tôi xin chân thành cảm ơn TS. Trần Thị Thúy Quỳnh, người đã dẫn dắt, định hướng cho nghiên cứu nhận dạng hệ thống với tri thức mới cho hệ thống truyền thông MIMO kích thước lớn. Tôi cũng xin cảm ơn các thầy, cô Khoa Điện tử - Viễn thông, các thành viên của Viện Tiên tiến về Kỹ thuật và Công nghệ, cùng các thầy, cô phản biện đã cho nhiều góp ý giúp tôi khắc phục những điểm còn hạn chế trong luận văn.

Nghiên cứu này nhận được sự tài trợ từ đề tài: “Kỹ thuật nhận dạng hệ thống: Từ xử lý mù truyền thống sang xử lý có tri thức mới” mã số 01/2019/TN của Quỹ phát triển khoa học và công nghệ Quốc gia (NAFOSTED).

Xin gửi lời cảm ơn đến Giáo sư Karim Abed-Meraim và các thành viên thuộc phòng thí nghiệm PRISME, Đại học Orleans, Cộng hoà Pháp đã tận tình giúp đỡ, hướng dẫn tôi trong quá trình tôi là thực tập sinh tại phòng thí nghiệm PRISME từ tháng 09/2022 đến 12/2022.

Cuối cùng, tôi xin được gửi lời cảm ơn đến gia đình, bạn bè đã luôn bên cạnh động viên, khích lệ, và giúp đỡ tôi trong suốt thời gian qua.

Trong quá trình thực hiện luận văn, do thời gian, kiến thức, và điều kiện nghiên cứu còn một số hạn chế nên không tránh khỏi thiếu sót, rất mong nhận được ý kiến đóng góp từ người đọc để tôi có thể tiếp tục phát triển hướng nghiên cứu này.

Hà Nội, ngày ... tháng ... năm 2023

Học viên

Đỗ Hải Sơn

TÓM TẮT

Tóm tắt:

Từ khóa:

MỤC LỤC

Lời cam đoan	
Lời cảm ơn	
Tóm tắt	
Mục lục	i
Danh mục ký hiệu và chữ viết tắt	iii
Danh mục hình vẽ	viii
Danh mục bảng biểu	ix
MỞ ĐẦU	1
CHƯƠNG 1. TỔNG QUAN VỀ CÁC PHƯƠNG PHÁP NHẬN DẠNG HỆ THỐNG TRONG TRUYỀN THÔNG KHÔNG DÂY	5
1.1. Mô hình hệ thống MIMO/mMIMO	6
1.2. Nhận dạng kênh không mù	7
1.2.1. Zero Forcing (ZF)	8
1.2.2. Minimum Mean Square Error (MMSE)	9
1.3. Nhận dạng kênh mù	9
1.4. Nhận dạng kênh bán mù	10
1.5. Nhận dạng kênh sử dụng học máy	11
CHƯƠNG 2. NHẬN DẠNG HỆ THỐNG SỬ DỤNG PHƯƠNG PHÁP BÁN MÙ	13
2.1. Sơ lược về thuật toán B-MRE	13
2.2. Đề xuất phương pháp nhận dạng hệ thống SB-MRE cho MIMO	16
2.3. Đề xuất giảm thiểu chi phí của thuật toán SB-MRE	18
2.4. Mô phỏng và đánh giá	18
2.5. Kết luận chương	21
CHƯƠNG 3. NHẬN DẠNG HỆ THỐNG SỬ DỤNG MẠNG HỌC SÂU	23
3.1. Giới thiệu về mạng nơ-ron sâu và mở rộng sâu (Deep unfolding)	23

3.2. Mạng nơ-ron học sâu DetNet	24
3.3. Đề xuất mạng nơ-ron sâu ISDNN cho nhận dạng kênh truyền	26
3.3.1. Bộ nhận dạng ISD cho hệ thống mMIMO	27
3.3.2. Đề xuất mạng nơ-ron sâu ISDNN	29
3.4. Mô phỏng và đánh giá	31
3.4.1. Tạo bộ dữ liệu	32
3.4.2. Đào tạo và đánh giá mô hình đề xuất	34
3.5. Kết luận chương	39
KẾT LUẬN	41
DANH MỤC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN VĂN	43
TÀI LIỆU THAM KHẢO	44
PHỤ LỤC	48
A. Thuật toán tối ưu Adam	48

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Danh mục ký hiệu		
STT	Ký hiệu	Giải thích
1	in thường	Vô hướng
2	in thường, đậm	Vector
3	in hoa, đậm	Ma trận
4	$(.)^\top$	Chuyển vị
5	$(.)^{-1}$	Nghịch đảo
6	$(.)^H$	Chuyển vị liên hợp phức
7	α	Tham số học của mạng ISDNN
8	δ	Tốc độ học / độ dài bước
9	\Im	Phần ảo
10	λ	Trọng số của thành phần B-MRE
11	$\mathbb{E}(\cdot)$	Công suất trung bình
12	\mathbf{e}	Véc-tơ phần dư / lỗi
13	\mathcal{L}	Hàm mất mát
14	\mathcal{O}	Độ phức tạp
15	\mathcal{U}	Phân bố đều
16	Ψ	Toán tử phi tuyến tính
17	Ψ_t	Hàm phi tuyến tính phân đoạn
18	\Re	Phần thực
19	σ	Độ lệch chuẩn
20	θ	Bộ tham số của việc học
21	$\ \cdot\ _F$	Frobenius norm
22	$\mathbf{0}_T$	Véc-tơ cột có kích thước $T \times 1$
23	arg	Vị trí
24	b	Bias của bộ biến đổi tuyến tính

25	\mathcal{CN}	Tạp âm AWGN phức
26	con	Phép nối
27	∂	Đạo hàm riêng
28	dB	Đơn vị decibel
29	diag	Thành phần đường chéo
30	\mathbf{G}	Ma trận của bộ nhận dạng tuyến tính
31	\mathbf{G}_H	Ma trận Gram của \mathbf{H}
32	\mathbf{H}	Ma trận kênh truyền
33	\mathbf{I}_K	Ma trận đơn vị kích thước K
34	K	Số lượng bộ cân bằng của B-MRE
35	L	Số ăng-ten thu
36	M	Bậc kênh truyền
37	min	Giá trị nhỏ nhất
38	N	Số lượng ký hiệu tại mỗi thời điểm bên thu lưu lại
39	N_p	Độ dài chuỗi ký hiệu pilot
40	N_s	Độ dài chuỗi ký hiệu s
41	ρ	Toán tử tuyến tính
42	\mathbf{s}	Các ký hiệu được gửi đi từ bộ phát
43	T	Số ăng-ten phát / người dùng bên phát
44	$\text{vec}(\mathbf{X})$	Véc-tơ hoá ma trận \mathbf{X}
45	\mathbf{w}	Tạp âm AWGN
46	w	Trọng số của bộ biến đổi tuyến tính
47	\mathbf{x}	Các ký hiệu bên thu nhận được

Danh mục chữ viết tắt			
STT	Chữ viết tắt	Giải thích tiếng Anh	Giải thích tiếng Việt
1	AI	Artificial intelligence	Trí tuệ nhân tạo
2	AWGN	Additive white Gaussian noise	Tạp âm trắng cộng sinh
3	B	Blind	Kỹ thuật nhận dạng mù
4	BER	Bit error rate	Tỷ lệ lỗi bit
5	C1	Constraint 1	Ràng buộc bậc hai
6	C2	Constraint 2	Ràng buộc tuyến tính
7	C3	Constraint 3	Ràng buộc công suất
8	CMA	Constant modulus algorithm	Thuật toán mô-đun không đổi
9	CSI	Channel state information	Thông tin về trạng thái kênh truyền
10	Data-driven	Data-driven	Hướng dữ liệu
11	DDCE	Decision-directed channel estimation	Phương pháp hướng quyết định
12	DetNet	Detection Network	Mạng nơ-ron học sâu phát hiện
13	DL	Deep learning	Học sâu
14	DNN	Deep-neural network	Mạng nơ-ron sâu
15	DoA	Direction of arrival	Hướng sóng đến
16	DoD	Direction of departure	Hướng phát sóng
17	FIR	Finite impulse response	Đáp ứng xung hữu hạn
18	GPR	Gaussian process regression	Thuật toán hồi quy Gaussian
19	HOS	Higer-order statistics	Đặc tính thống kê bậc cao
20	i.i.d	Independent and identically distributed	Biến độc lập và phân phối đồng nhất
21	ICA	Independent component analysis	Phân tích thành phần độc lập
22	ISD	Iterative sequential detection	Mạng học sâu phát hiện tuần tự lặp lại
23	ISDNN	Iterative sequential Deep-neural network	Mạng nơ-ron sâu tuần tự lặp lại
24	LMMSE	Linear Minimum mean square error	Kỹ thuật ước lượng lỗi bình phương trung bình nhỏ nhất tuyến tính

25	LMS	Least mean squares	Kỹ thuật trung bình bình phương nhỏ nhất
26	LS	Least square	Kỹ thuật bình phương nhỏ nhất
27	LSTM	Long/short-term memory	Mạng trí nhớ dài hạn/ngắn hạn
28	MIMO	Multi-input multi-output	Hệ thống đa đầu vào đa đầu ra
29	ML	Machine learning	Học máy
30	MLE	Maximum likelihood estimator	Bộ ước lượng hợp lý cực đại
31	mMIMO	Massive Multi-input multi-output	Hệ thống đa đầu vào đa đầu ra kích thước lớn
32	MMSE	Minimum mean square error	Kỹ thuật lỗi trung bình phương sai tối thiểu
33	mmWave	Millimeter wave	Bước sóng mi-li-mét
34	Model-driven	Model-driven	Hướng mô hình
35	MRE	Mutually referenced equalizers	Bộ cân bằng kênh tham chiếu
36	NB	Non-blind	Kỹ thuật nhận dạng không mù
37	NN	Neural network	Mạng nơ-ron
38	PC	Pilot contamination	Ô nhiễm pilot
39	PCA	Principal components analysis	Phân tích thành phần chính
40	Pilot	Pilot	Ký hiệu hoa tiêu
41	Pilot-assisted	Pilot-assisted	Phương pháp dựa trên tín hiệu hoa tiêu
42	PSK	Phase-shift keying	Điều chế pha
43	QAM	Quadrature amplitude modulation	Điều chế biên độ cầu phương
44	rc	Reduced cost	Giảm thiểu chi phí
45	RL	Reinforcement learning	Học tăng cường
46	RLS	Recursive least squares	Kỹ thuật bình phương đệ quy nhỏ nhất
47	SB	Semi-blind	Kỹ thuật nhận dạng bán mù
48	SER	Symbol error rate	Tỷ lệ sai số ký hiệu
49	SIMO	Single-input multi-output	Hệ thống đơn đầu vào đa đầu ra
50	SISO	Single-input single-output	Hệ thống đơn đầu đơn vào đầu ra
51	SNR	Signal noise ratio	Tỷ lệ tín hiệu tạp âm
52	SOS	Second-order statistics	Đặc tính thống kê bậc hai

53	Training-based	Training-based	Phương pháp dựa trên việc đào tạo
54	ZF	Zero forcing	Kỹ thuật ép không

DANH MỤC HÌNH VẼ

Hình 1.1	Phân loại các phương pháp ước lượng kênh truyền viễn thông.	5
Hình 1.2	Mô hình minh hoạ hệ thống truyền thông MIMO.	6
Hình 2.1	Minh hoạ hệ thống SIMO sử dụng phương pháp nhận dạng B-MRE.	14
Hình 2.2	Minh hoạ chuỗi một ký hiệu s_t sử dụng cho SB-MRE.	16
Hình 2.3	So sánh SER của thuật toán đề xuất SB-MRE với các phương pháp nhận dạng tuyến tính.	20
Hình 2.4	So sánh SER của thuật toán đề xuất SB-MRE với các số lượng pilot (N_p) và SNR khác nhau.	21
Hình 2.5	So sánh SER của thuật toán đề xuất SB-MRE với trọng số thành phần B-MRE (λ) và SNR khác nhau.	22
Hình 3.1	Minh hoạ sử dụng DNN để nhận dạng hệ thống viễn thông.	23
Hình 3.2	Kiến trúc của một lớp trong mô hình mạng DetNet [32].	26
Hình 3.3	Hàm phi tuyến tính phân đoạn $\psi_{t_k}(s)$ được sử dụng trong DetNet.	27
Hình 3.4	Kiến trúc của một lớp trong mô hình mạng nơ-ron sâu ISDNN đề xuất.	31
Hình 3.5	Minh hoạ một số hàm kích hoạt được dùng trong mô hình đề xuất.	32
Hình 3.6	Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM.	33
Hình 3.7	Phân bố khi gieo ngẫu nhiên của các hệ số $h_{l,t}$ trong ma trận \mathbf{H}	34
Hình 3.8	Sự hội tụ của quá trình đào tạo mô hình ISDNN và DetNet.	36
Hình 3.9	Độ chính xác của mô hình ISDNN so sánh với DetNet và các bộ nhận dạng tuyến tính.	37
Hình 3.10	Sự hội tụ của quá trình đào tạo các mô hình ISDNN với các sai số kênh truyền đầu vào khác nhau.	38
Hình 3.11	Độ chính xác của mô hình ISDNN với các sai số kênh truyền đầu vào khác nhau.	39

DANH MỤC BẢNG BIỂU

Bảng 2.1	Các tham số mô phỏng hệ thống truyền thông không dây của SB-MRE.	18
Bảng 3.1	Các tham số mô phỏng hệ thống truyền thông không dây của mạng nơ-ron sâu ISDNN được đề xuất.	32
Bảng 3.2	Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM.	33
Bảng 3.3	So sánh độ phức tạp của các thuật toán nhận dạng kênh truyền.	36

MỞ ĐẦU

Lý do chọn đề tài

Theo [2], trong năm 2021, chỉ riêng các thiết bị di động đã đòi hỏi tổng lưu lượng truy cập internet đạt 48,27 Petabytes/tháng, và tăng trưởng 46% hàng năm. Tốc độ truy cập của các thế hệ mạng di động phổ biến ở Việt Nam như 4G là trung bình trên 28,2 Megabits/giây và tăng lên từ 29 đến 47% hàng năm. Với sự ra trường nhanh chóng của số lượng các thiết bị di động và đòi hỏi về chất lượng nội dung của người dùng khiến yêu cầu về truyền tải nhanh và hiệu quả trong các hệ thống truyền thông không dây luôn là chủ đề nghiên cứu được quan tâm.

Tuy nhiên, việc truyền tải không dây luôn gặp phải một hạn chế cố hữu đó là ảnh hưởng bởi kênh truyền vô tuyến là biến dạng tín hiệu. “**Nhận dạng hệ thống truyền thông**” được hiểu là ước lượng sự ảnh hưởng của kênh truyền vô tuyến đến việc truyền nhận tín hiệu, gọi ngắn gọn là “ước lượng kênh truyền”. Ngay từ các thế hệ mạng di động đầu tiên như 2G [40], các chuỗi tín hiệu hoa tiêu (pilot sequence) được biết trước ở cả bên phát và thu đã được sử dụng để ước lượng sự ảnh hưởng của kênh truyền và khôi phục dạng tín hiệu ở bên thu (NB - Non-blind) [20]. Đến thế hệ mạng di động 4G, 5G, và cao hơn, việc truyền tải đơn ăng-ten (antenna) đã được thay thế bằng các hệ thống đa đầu vào đa đầu ra (MIMO - Multi-input multi-output) hay lớn hơn nữa là MIMO kích thước lớn (mMIMO - Massive MIMO). Điều này khiến việc ước lượng kênh truyền trong các hệ thống MIMO, mMIMO trở nên phức tạp, yêu cầu các chuỗi pilot dài hơn [22], dẫn đến hiệu quả về mặt phổ thời gian tần số của việc truyền tải bị giảm đi.

Nhiều phương pháp nghiên cứu đã được đề xuất để giảm thiểu số lượng pilot cần thiết cho việc ước lượng kênh truyền. Mà tiêu biểu trong số đó là ba hướng tiếp cận:

1. Các thuật toán nhận dạng “mù” (B - Blind) ở đây được hiểu là khi xử lý (nhận dạng), bộ xử lý “không nhìn thấy” (không có thông tin) của đầu vào. Các thuật toán xử lý tín hiệu mù phát triển mạnh trong thập kỷ 90 [1], tuy nhiên các phương pháp xử lý mù thường yêu cầu các thông số thống kê của tín hiệu mà thông thường không biết trước trong các hệ thống truyền thông thực, hơn nữa, độ chính xác mà các thuật toán này đưa ra cũng thấp hơn đáng kể khi so sánh với các phương pháp sử dụng pilot truyền thống. Do vậy, các thuật toán mù cũng ít được quan tâm trong những thế hệ mạng viễn thông di động trước 5G.

2. Các thuật toán nhận dạng “bán mù” (SB - Semi-blind) là phương pháp cải tiến của B và được quan tâm trong các năm gần đây [16, 17, 34]. Đây là kỹ thuật kết hợp các thông tin từ hướng tiếp cận mù truyền thống và các dạng thông tin khác, ví dụ [29]: số lượng nhỏ pilot, hướng sóng đến (DoA - Direction of arrival), tọa độ người dùng, ... Điều này giúp giảm đi số lượng pilot cần thiết cho việc nhận dạng hệ thống nhằm tăng hiệu quả phổ nhưng vẫn giữ được độ chính xác cần thiết, giảm độ phức tạp, và cho khả năng ứng dụng rộng rãi hơn.
3. Các thuật toán nhận dạng sử dụng học máy, học sâu (ML - Machine learning; DL - Deep learning) cũng là lĩnh vực nghiên cứu dành được nhiều sự quan tâm [48] trong các năm gần đây. Ưu điểm của các kỹ thuật sử dụng ML, DL là tính đa dạng, khi hướng tiếp cận ML, DL sử dụng cho mục đích xử lý các loại tín hiệu như hình ảnh, âm thanh đã đạt được các bước tiến rõ rệt. Đầu vào của các mạng DL được sử dụng để nhận dạng hệ thống rất linh hoạt, có thể tương ứng với cả ba hướng tiếp cận: pilot, mù, và bán mù kể trên. Sau quá trình huấn luyện, các mô hình (model) học máy có thể hoạt động độc lập như một bộ cân bằng mù/bán mù, khi chỉ cần đưa các tín hiệu thu được đi qua model, và các tín hiệu được khôi phục sẽ được trả về mà không cần đến các chuỗi pilot, hay thông tin về trạng thái kênh truyền (CSI - Channel state information).

Có thể thấy, với SB và DL, thông tin về kênh truyền cần thiết phải được biết trước đó. Thay vào đó, các phương pháp này có thể sử dụng các tín hiệu mẫu được thu thập để đưa ra ước lượng chính xác của thông tin kênh truyền, giảm thiểu sự phụ thuộc vào kiến thức chính xác về kênh truyền. Vì vậy, SB và DL là các phương pháp tiên tiến và hứa hẹn trong việc ước lượng kênh truyền trong các hệ thống truyền thông hiện đại.

Từ thực tế và những phân tích nêu trên, luận văn tập trung nghiên cứu hai mục tiêu sau phương pháp nhận dạng hệ thống bán mù, học máy, học sâu và đề xuất các giải thuật mới để cải thiện các phương pháp nhận dạng này, bao gồm:

1. Phát triển thuật toán bán mù dựa trên một phương pháp nhận dạng mù truyền thống cho các hệ thống MIMO và mMIMO.
2. Phát triển một phương pháp nhận dạng sử dụng DL cho các hệ thống mMIMO.

Phương pháp nghiên cứu

Trong luận văn, để đạt được mục tiêu nghiên cứu học viên đã tìm hiểu các tài liệu, bài báo, tạp chí quốc tế, ... có uy tín, thực hiện việc tính toán mô hình dữ liệu, phân tích

số học để đưa ra các hướng giải quyết hợp lý, và sau đó kiểm nghiệm lại kết quả bằng hình thức mô phỏng trên Matlab, Python.

Cụ thể để phát triển các thuật toán mới như mục tiêu đề ra các bước nghiên cứu sau đã được thực hiện trong luận văn:

1. Tìm hiểu về tổng quan các thuật toán nhận dạng hệ thống trong truyền thông không dây với các hướng tiếp cận khác nhau.
2. Xác định bài toán cụ thể và mục tiêu của nghiên cứu.
3. Lựa chọn thuật toán bán truyền thống phù hợp để phát triển lên SB và một phương pháp sử dụng học sâu để nhận dạng kênh truyền có thể cải tiến.
4. Xây dựng mô hình toán học, huấn luyện mô hình, và tối ưu hóa các thuật toán đề xuất.
5. Đánh giá và phân tích kết quả ở nhiều kịch bản khác nhau, so sánh với các nghiên cứu đi trước.

Nội dung nghiên cứu

- Tìm hiểu về các thuật toán ước lượng kênh truyền trong những năm gần đây.
- Xây dựng mô hình toán học hệ thống MIMO/mMIMO để mô tả quá trình truyền tải tín hiệu qua kênh truyền.
- Phát triển thuật toán SB dựa trên phương pháp bộ cân bằng kênh tham chiếu (MRE - Mutually referenced equalizers).
- Phát triển một mạng nơ-ron học sâu dựa trên giải thuật phát hiện tuần tự lặp lại (ISD - Iterative sequential detection) để ước lượng kênh truyền viễn thông.
- Đánh giá, kiểm tra các thuật toán được đề xuất trên các công cụ mô phỏng Matlab, Python.

Đóng góp của đề tài

Với sự hiểu biết của học viên, những kết quả nghiên cứu trong luận văn đã đạt được mục đích nghiên cứu đề ra. Những kết quả này bao gồm:

- Tổng quan về các thuật toán nhận dạng hệ thống truyền thông MIMO kích thước lớn.

- Đề xuất thuật toán SB-MRE cho ước lượng kênh truyền.
- Đề xuất một mạng học sâu ISDNN cho việc ước lượng kênh truyền của hệ thống mMIMO.

Kết quả nghiên cứu trong khuôn khổ luận văn đã được công bố trên ...

1.

Bố cục của luận văn

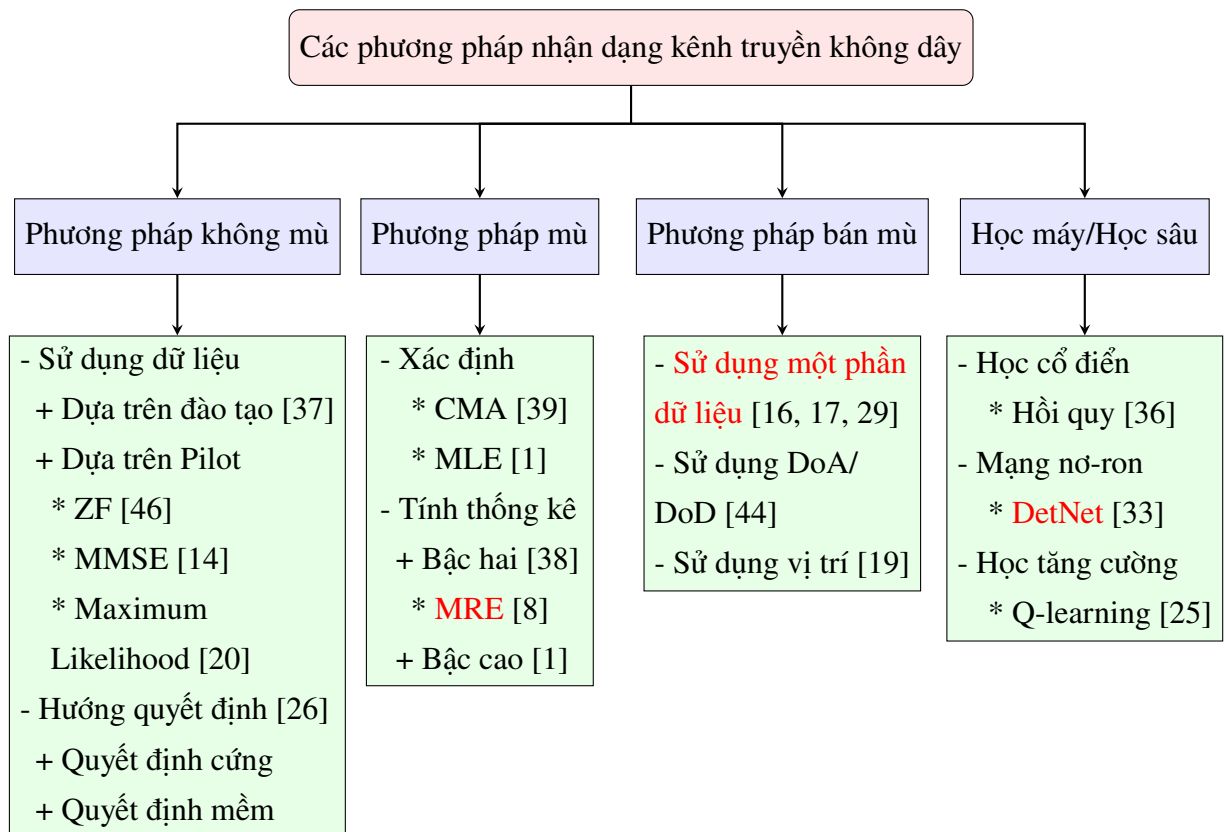
Nội dung chính của luận văn được trình bày như sau:

- Mở đầu: Trình bày mục đích, phương pháp nghiên cứu, nội dung, đóng góp và bố cục của luận văn.
- Chương 1: Trình bày mô hình hệ thống viễn thông MIMO/mMIMO và tổng quan về các phương pháp nhận dạng hệ thống trong truyền thông không dây.
- Chương 2: Trình bày sơ lược về thuật toán B-MRE gốc, sau đó đề xuất phương pháp SB-MRE cho MIMO/mMIMO. Kiểm nghiệm và đánh giá kết quả thông qua mô phỏng Matlab.
- Chương 3: Trình bày sơ lược về mạng DetNet (Detection network) phục vụ cho ước lượng kênh truyền viễn thông. Đề xuất một mô hình mạng ISDNN nhằm giảm độ phức tạp cho mô hình đầu ra. Tạo bộ dữ liệu, đào tạo và đánh giá kết quả thông qua mô phỏng Python.
- Kết luận: Đưa ra kết luận về hai thuật toán được đề xuất và đề xuất các giải pháp để cải thiện hệ thống.

CHƯƠNG 1

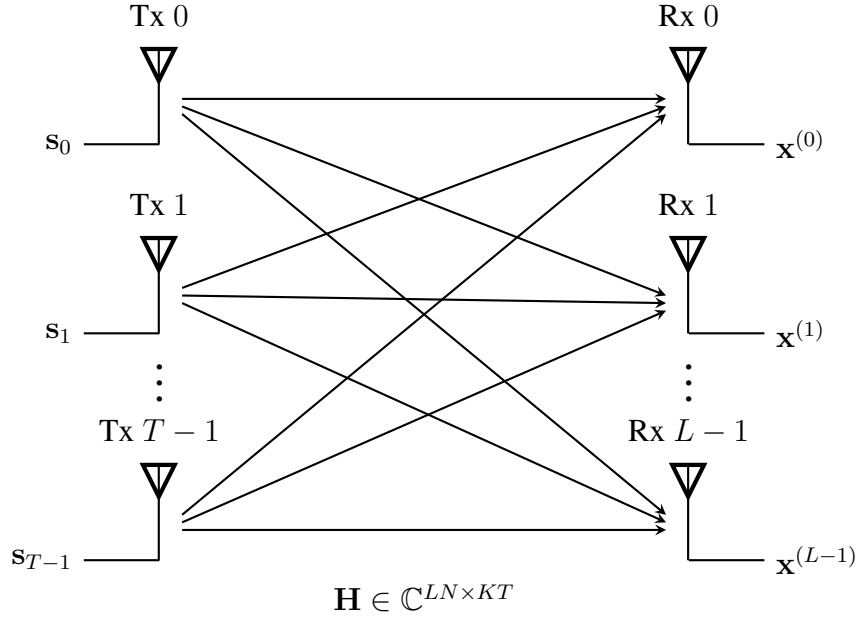
TỔNG QUAN VỀ CÁC PHƯƠNG PHÁP NHẬN DẠNG HỆ THỐNG TRONG TRUYỀN THÔNG KHÔNG DÂY

Việc nhận dạng hệ thống trong truyền thông không dây đã luôn được phát triển ngay từ những thế hệ mạng không dây đầu tiên [40]. Ngày nay, các thuật toán ước lượng kênh truyền không dây đã đạt được các bước tiến đáng kể về độ chính xác và dựa trên đặc điểm của các thuật toán có thể chia thành bốn hướng tiếp cận chính như trên hình 1.1. Bao gồm phương pháp không mù (Non-blind), mù (B), bán mù (SB), và dựa trên học máy, học sâu (AI-based) [42]. Với mỗi phương pháp, rất nhiều thuật toán đã được đề xuất và cho hiệu quả trong các tình huống cụ thể như các trích dẫn trên hình 1.1. Từ cách phân loại kể trên, chương đầu tiên sẽ đưa ra mô hình hệ thống MIMO/mMIMO sử dụng cho luận văn, sau đó sẽ trình bày đôi nét cơ bản về các phương pháp ước lượng kênh truyền này, trong số đó, một số thuật toán được dùng để so sánh kết quả trong các chương sau của luận văn sẽ được trình bày chi tiết.



Hình 1.1. Phân loại các phương pháp ước lượng kênh truyền viễn thông.

1.1. Mô hình hệ thống MIMO/mMIMO



Hình 1.2. Mô hình minh hoạ hệ thống truyền thông MIMO.

Hình 1.2 minh hoạ một hệ thống thu phát MIMO/mMIMO, với T ăng-ten phát và L ăng-ten thu. Mỗi kênh truyền vô tuyến giữa ăng-ten phát thứ t và ăng-ten nhận thứ l sẽ được mô hình hoá bằng một bộ lọc đáp ứng xung hữu hạn (FIR - Finite impulse response) tương ứng là một véc-tơ của các hệ số bộ lọc có độ dài $M + 1$. Giả sử ở mỗi thời điểm n , mỗi ăng-ten thu sẽ thu thập một chuỗi N ký hiệu (symbol) liên tiếp. Từ các giả thiết trên, mô hình toán học của hệ thu phát MIMO được biểu diễn dưới dạng ma trận như sau

$$\mathbf{x}(n) = \sum_{t=0}^{T-1} \mathbf{H}_t \mathbf{s}_t(n) + \mathbf{w}_t \quad (1.1)$$

trong đó $\mathbf{s}_t(n) \in \mathbb{C}^{M+N \times 1}$ là các ký hiệu được gửi đi từ bộ phát thứ t . Ma trận kênh truyền dạng tích chập (convolution) [8] giữa bộ phát thứ t và toàn bộ L ăng-ten thu được ký hiệu \mathbf{H}_t . Giả sử rằng $\mathbf{H}_t \in \mathbb{C}^{LN \times K}$ là một ma trận đầy đủ hạng theo cột (full column rank) với hạng $K = M + N$. Tiếp đến, $\mathbf{x}(n) \in \mathbb{C}^{LN \times 1}$ là véc-tơ biểu diễn tín hiệu thu được từ L ăng-ten. Cuối cùng, $\mathbf{w}_t \in \mathbb{C}^{LN \times 1}$ đại diện cho tạp âm trắng cộng sinh (AWGN - Additive white Gaussian noise). Giả sử kênh truyền và tạp âm ở các kênh khác nhau là độc lập và phân bố giống nhau (i.i.d) với các phân bố lần lượt được chọn là $\mathcal{CN}(0, \sigma_{\mathbf{H}_t}^2 \mathbf{I})$ và $\mathcal{CN}(0, \sigma^2 \mathbf{I})$. Dưới đây là biểu diễn đầy đủ của các thành phần kể trên với $(\cdot)^\top$ là phép chuyển vị (transpose) ma trận.

$$\mathbf{s}_t(n) = [s_t(n), s_t(n-1), \dots, s_t(n-K+1)]^\top \quad (1.2)$$

$$\mathbf{H}_t = \begin{pmatrix} h_{t,0}^{(0)} & \cdots & h_{t,M}^{(0)} & \overset{\leftarrow K}{\rightarrow} 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & h_{t,0}^{(0)} & \cdots & h_{t,M}^{(0)} \\ \vdots & \cdots & \vdots & \cdots & \cdots & \vdots \\ h_{t,0}^{(L-1)} & \cdots & h_{t,M}^{(L-1)} & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & h_{t,0}^{(L-1)} & \cdots & h_{t,M}^{(L-1)} \end{pmatrix} \uparrow_{LN} \quad (1.3)$$

$$\mathbf{x}(n) = [x^{(0)}(n), \dots, x^{(0)}(n - N + 1), \dots, x^{(L-1)}(n), \dots, x^{(L-1)}(n - N + 1)]^\top \quad (1.4)$$

Cần lưu ý rằng, cách biểu diễn toán học dưới dạng ma trận số phức như trên chỉ phù hợp trên lý thuyết và các phần mềm mô phỏng như Matlab. Các phương pháp sử dụng ML/DL nhằm nhận dạng kênh thường được phát triển trên ngôn ngữ Python và các thư viện nền tảng thông dụng như Tensorflow¹ của Google hay Pytorch² của Facebook, cả hai thư viện này không trực tiếp hỗ trợ các phép toán/toán tử với số phức. Do vậy, không làm mất đi tính tổng quát, trong chương 3, các ma trận trên mô hình kênh kể trên sẽ được biểu diễn dưới dạng ma trận của các thành phần thực (\Re) và ảo (\Im) riêng biệt như sau

$$\mathbf{s}_t = \begin{bmatrix} \Re(\mathbf{s}_t) \\ \Im(\mathbf{s}_t) \end{bmatrix}; \mathbf{x} = \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}; \mathbf{w}_t = \begin{bmatrix} \Re(\mathbf{w}_t) \\ \Im(\mathbf{w}_t) \end{bmatrix} \quad (1.5)$$

$$\mathbf{H}_t = \begin{bmatrix} \Re(\mathbf{H}_t) & -\Im(\mathbf{H}_t) \\ \Im(\mathbf{H}_t) & \Re(\mathbf{H}_t) \end{bmatrix} \quad (1.6)$$

1.2. Nhận dạng kênh không mù

Như trên hình 1.1, các phương pháp nhận dạng kênh không mù có thể chia làm hai nhóm chính, bao gồm các phương pháp sử dụng dữ liệu (Data-aided) [42] và các phương pháp dựa trên hướng quyết định (Decision-directed) [26]. Các thuật toán sử dụng dữ liệu có thể chia làm hai loại nhỏ hơn, gồm có các phương pháp dựa trên việc đào tạo (Training-based) và các phương pháp dựa trên tín hiệu hoa tiêu (Pilot-assisted). Khác biệt chính giữa hai phương pháp là loại tín hiệu được dùng để ước lượng kênh truyền. Với Training-based, bên phát sẽ phát các dữ liệu huấn luyện gốc, bên thu chỉ biết thời điểm dữ liệu huấn luyện này được truyền nhưng không biết trước thông tin của dữ liệu.

¹<https://github.com/tensorflow/tensorflow>

²<https://github.com/pytorch/pytorch>

Tín hiệu bên thu nhận được gồm tín hiệu gốc và tín hiệu đã bị méo, từ đó, một mô hình ước lượng được huấn luyện bằng cách tối ưu hóa một hàm mất mát (loss function) giữa kết quả ước lượng kênh truyền và giá trị thực tế của kênh truyền. Với Pilot-assisted, các ký hiệu pilot được chèn trực tiếp vào khung dữ liệu gửi đi, và bên phát biết cả thời gian, vị trí, và giá trị gốc của các ký hiệu pilot này. Từ đó, bên thu có thể ước lượng ra ảnh hưởng của kênh truyền đến các tín hiệu pilot và nội suy ra ảnh hưởng của kênh truyền đến toàn bộ khung dữ liệu. Các giải thuật phổ biến được sử dụng cho phương pháp Data-aided có thể kể đến như bộ phát hiện ép về không (ZF - Zero forcing), lỗi trung bình phương sai tối thiểu (MMSE - Minimum mean square error) [14]. Hai giải thuật này là các giải thuật tuyến tính và sẽ được trình bày chi tiết hơn ở mục 1.2.1 và 1.2.2. Tuy phổ biến và được áp dụng trong các hệ truyền thông thực tế, nhưng các phương pháp sử dụng dữ liệu để ước lượng kênh truyền có một nhược điểm đó là giảm hiệu quả sử dụng phổ do một phần băng thông bị lãng phí để truyền tải các dữ liệu huấn luyện hoặc pilot.

Phương pháp hướng quyết định (DDCE - Decision-directed channel estimation) cũng dựa trên việc sử dụng dữ liệu, tuy nhiên, thay vì ước lượng kênh truyền chỉ trong một bước DDCE có thêm một bước nữa [42]. Cụ thể, tại bước một, DDCE vẫn ước lượng kênh truyền dựa trên một trong hai phương pháp Training-based hoặc Pilot-assisted như Data-aided. Sau đó, khôi phục các tín hiệu dựa trên trạng thái kênh truyền vừa ước lượng được. Ở bước tiếp theo, các dữ liệu mới được khôi phục sẽ tiếp tục được đưa vào thuật toán ước lượng nhằm cập nhật trạng thái thông tin về kênh truyền cho đến khi các ký hiệu trong một phiên được truyền hết. Chi tiết hơn, bộ thu sẽ so sánh ký tự đã nhận được với ký tự được dự đoán dựa trên ký tự trước đó và ước lượng kênh truyền hiện tại. Nếu có sai sót giữa ký tự đã nhận và ký tự dự đoán, bộ thu sẽ điều chỉnh ước lượng kênh truyền để cải thiện độ chính xác của dự đoán ký tự tiếp theo. Quá trình này được lặp lại cho mỗi ký tự nhận được. Vậy ra quyết định bit là 0 hay 1 trong DDCE sẽ được thực hiện theo hai giải thuật gồm quyết định mềm (soft) và quyết định cứng (hard). Quyết định mềm [28] sẽ xác định giá trị của các bit dữ liệu bằng cách tính toán xác suất bit đó được truyền qua kênh truyền. Ngược lại, với quyết định cứng [35], một ngưỡng xác định được đưa ra, nếu lớn hơn ngưỡng này sẽ là bit 1, ngược lại là bit 0. Tuy nhiên, phương pháp DDCE có điểm hạn chế đó là quá trình ước lượng lại bị phụ thuộc vào các dữ liệu cũ, dẫn đến việc, có thể kênh truyền hiện tại không còn tương ứng với các dữ liệu từ thời điểm quá khứ. Điều này dẫn đến các lỗi tích lũy và làm giảm hiệu năng của hệ thống nhận dạng.

1.2.1. Zero Forcing (ZF)

Các giải thuật nhận dạng tuyến tính dựa trên các phép biến đổi tuyến tính của các tín hiệu nhận được x . Các giải thuật này thường có độ phức tạp thấp, hoặc trung bình.

Nhưng độ phức tạp sẽ tăng lên nếu hệ thống có số chiều lớn, ví dụ số lượng ăng-ten T hay L rất lớn trong mMIMO dẫn đến phép nghịch đảo ma trận tiêu tốn nhiều tài nguyên tính toán hơn. Một bộ nhận dạng tuyến tính có thể biểu diễn như bên dưới

$$\mathbf{s} = \mathbf{G}\mathbf{x} \quad (1.7)$$

ZF là phương pháp đơn giản nhất của các thuật toán nhận dạng tuyến tính. Trong đó, ma trận kênh truyền \mathbf{H} sẽ được nghịch đảo để loại bỏ ảnh hưởng của kênh truyền. Ma trận làm bằng (equalizer matrix) \mathbf{G}_{ZF} của bộ nhận dạng ZF như sau

$$\mathbf{G}_{ZF} = \left(\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \quad (1.8)$$

trong đó $(.)^H$ là ma trận Hermitian. Với \mathbf{G}_{ZF} , tín hiệu gốc được khôi phục/ước lượng bằng cách

$$\hat{\mathbf{s}}_{ZF} = \left(\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \mathbf{x} \quad (1.9)$$

1.2.2. Minimum Mean Square Error (MMSE)

Hiệu năng của bộ nhận dạng ZF thường bị ảnh hưởng bởi nhiễu AWGN. Do vậy, MMSE là giải thuật tốt hơn bằng việc kết hợp thêm thông tin phương sai của nhiễu trước khi nghịch đảo ma trận. Ma trận làm bằng \mathbf{G}_{MMSE} của bộ nhận dạng MMSE được biểu diễn dưới dạng

$$\mathbf{G}_{MMSE} = \left(\mathbf{H}^H \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I} \right)^{-1} \mathbf{H}^H \quad (1.10)$$

với σ^2 là phương sai của nhiễu AWGN, $\mathbb{E}(\mathbf{s})$ là công suất trung bình của mỗi ký hiệu gửi đi, và \mathbf{I} là ma trận đơn vị. với \mathbf{G}_{MMSE} , tín hiệu gốc được khôi phục như sau

$$\hat{\mathbf{s}}_{MMSE} = \left(\mathbf{H}^H \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I} \right)^{-1} \mathbf{H}^H \mathbf{x} \quad (1.11)$$

Ưu điểm của MMSE, các giá trị thấp trong quá trình đảo ngược có thể dẫn đến hiện tượng khuếch đại tạp âm (deep null) khi sử dụng ZF, được khắc phục bởi công suất tạp âm khác không. Tuy nhiên, có thể nhận thấy cả hai giải thuật ZF và MMSE cần các chuỗi pilot để ước lượng kênh truyền, sau đó nội suy ra ma trận \mathbf{H} .

1.3. Nhận dạng kênh mù

Các kỹ thuật nhận dạng hệ thống mù (hoặc tương tự như giải mã mù hoặc cân bằng mù) đã được biết đến từ đầu những năm 1980. Theo [42], có thể chia các thuật toán mù

vào hai nhóm chính. Thứ nhất là các kỹ thuật ước lượng kênh truyền dựa trên đặc tính thống kê của tín hiệu thu được, có thể là đặc tính thống kê bậc hai (SOS - Second-order statistics), hoặc bậc cao (HOS - Higher-order statistics). Cách tiếp cận SOS được đề xuất trong [38] yêu cầu các tín hiệu có đặc tính chu kỳ hoặc đa dạng kênh (channel diversity) với các hệ thống đơn đầu đơn vào đầu ra (SISO - Single-input single-output). SOS có ưu điểm là yêu cầu lượng dữ liệu ít hơn để có được các ước tính thống kê đáng tin cậy tương đương với phương pháp HOS. Tuy nhiên, HOS [10] có lợi thế là cung cấp thông tin về pha mà không cần đa dạng kênh với đánh đổi là cần một lượng lớn dữ liệu lấy mẫu và khả năng tính toán cao hơn. Trong luận văn, thuật toán mù khai thác tính thống kê bậc hai là MRE sẽ được chọn để phát triển thành SB. Thuật toán MRE [8] khai thác đặc trưng tham chiếu của hệ thống gồm nhiều cảm biến thu (sensor) hay được hiểu là đa ăng-ten trong một hệ đơn đầu vào đa đầu ra (SIMO - Single-input multi-output). Nhóm tác giả đề xuất phương pháp ước lượng một tập N bộ lọc để làm bằng kênh, tương ứng với N mẫu độ trễ ở các ăng-ten khác nhau.

Nhóm kỹ thuật thứ hai đó là khai thác các thông tin đã xác định (deterministic) của tín hiệu hoặc hệ thống. Có thể kể đến như thuật toán mô-đun không đổi (CMA - Constant modulus algorithm) [39] khai thác đặc trưng là giá trị mô-đun không đổi của các tín hiệu phức khi sử dụng các bộ điều chế như điều chế pha (PSK - Phase-shift keying), điều chế biên độ cầu phương 4 điểm (4-QAM - Quadrature amplitude modulation). Từ đó, nhóm tác giả đề xuất cân bằng kênh truyền bằng một bộ lọc thích nghi (adaptive filter) để đưa mô-đun của tín hiệu thu được về các giá trị chuẩn của PSK hay 4-QAM. Trong [3], nhóm tác giả sử dụng đặc trưng thưa (sparsity) của tín hiệu thường xuất hiện nhiều trong các kênh truyền mMIMO hay bước sóng mi-li-mét (mmWave - Millimeter wave) hiện nay. Bằng cách sử dụng tính chất thưa, các tín hiệu gốc có thể được khôi phục trong trường hợp hệ thống dưới mức xác định (underdetermined). Trong một số điều kiện cụ thể, việc áp dụng ràng buộc thưa có thể làm cải thiện hiệu năng của việc nhận dạng hệ thống mù.

1.4. Nhận dạng kênh bán mù

Các phương pháp nhận dạng kênh bán mù được sinh từ sự kết hợp của các kỹ thuật Non-blind (NB) và Blind. Giải pháp lai (hybrid) này được kỳ vọng sẽ giảm thiểu lượng pilot cần thiết mà vẫn bù đắp lại được độ chính xác bằng các thông tin từ kỹ thuật mù mang lại. Các tiếp cận đơn giản nhất đó là kết hợp trực tiếp các bộ nhận dạng như ZF, MMSE với các thông tin thống kê SOS, HOS đã được trình bày ở trên. Các công bố [16, 29, 43] đi theo hướng tiếp cận này đều cho ra các kết quả vượt trội khi so với với NB truyền thống trong một số điều kiện nhất định. Ngoài ra, việc kết hợp các thông tin xác định của các bộ cân bằng mù như được trình bày ở mục 1.3 cũng là các hướng

ngiên cứu tiềm năng trong tương lai.

Tiếp đến, ngoài các đặc trưng của tín hiệu, các thông tin bên lề (side-information) của hệ thống thu phát cũng có thể được xem xét để cải thiện khả năng nhận dạng kênh truyền. Có thể kể đến như sử dụng thêm thông tin hướng sóng đến/đi (DoA/DoD) như trong [44], nhóm tác giả đã đề xuất sử dụng DoA của các người dùng khác nhau để giảm thiểu/loại bỏ sự ảnh hưởng của ô nhiễm pilot (PC - Pilot contamination) qua đó hiệu suất của việc nhận dạng kênh truyền đã được cải thiện. Tiếp đến, [19] đề xuất sử dụng thông tin về tọa độ/vị trí (location) người dùng để đánh giá đáp ứng tần số kênh truyền mmWave. Kết quả mô phỏng cho thấy cả độ chính xác và độ phức tạp của mô hình ước lượng đều giảm đi khi có thêm loại thông tin bên lề này.

1.5. Nhận dạng kênh sử dụng học máy

Nhận dạng kênh truyền sử dụng ML/DL là hướng tiếp cận bùng nổ trong các năm trở lại đây. Do các bước tiến đã đạt được trước đó của việc xử lý các loại tín hiệu âm thanh, hình ảnh sử dụng các mạng học sâu. Việc chuyển tiếp các kỹ thuật sẵn có này sang viễn thông được xảy ra nhanh chóng và bước đầu các nghiên cứu đã chỉ ra các kết quả tiềm năng. Điểm khác biệt của hướng tiếp cận này đó là nó bao hàm được lý thuyết của cả ba hướng tiếp cận kể trên bao gồm mù, bán mù, và không mù. Tuy nhiên, thay vì việc tìm các phương pháp tối ưu và nghiệm chính xác, ML/DL sử dụng các thuật toán ML cơ bản, mạng nơ-ron (NN - Neural network), hay học tăng cường (RL - Reinforcement learning) cho các đầu vào của hệ thống nhận dạng B, SB, NB.

Các phương pháp sử dụng học cổ điển để nhận dạng kênh truyền được phát triển trước tiên, do độ phức tạp ở mức thấp. Trong [36], việc ước lượng ma trận làm bằng G_{MMSE} được thay thế bằng thuật toán hồi quy Gaussian (GPR - Gaussian process regression). Các ưu điểm của GPR như (i) tỷ lệ lỗi bit (BER - Bit error rate) thấp hơn MMSE truyền thống; (ii) nội suy chính xác hơn ước tính kênh ở giữa các ký hiệu pilot so với kỹ thuật nội suy tuyến tính. Ngoài phương pháp hồi quy, các giải thuật cổ điển của học máy như giảm số chiều của dữ liệu (PCA - Principal components analysis, ICA - Independent component analysis), học Bayesian cũng được đề xuất và cho thấy sự hiệu quả [42].

Các phương pháp nhận dạng sử dụng các mạng nơ-ron còn có những bước tiến rõ ràng hơn, khi NN phức tạp hơn và số lượng tham số đào tạo cũng là rất lớn để đáp ứng được các mô hình kênh phức tạp. Các nghiên cứu trong mục 1.4 như [19, 43] cũng sử dụng các thông tin bên lề cho SB nhưng thay vì phương pháp tối ưu đại số, các mạng nơ-ron sâu (DNN - Deep-neural network) đã được đề xuất để ước lượng kênh truyền.

Một trong những mạng DNN đầu tiên được đề xuất cho việc nhận dạng hệ thống MIMO đó là mạng phát hiện (DetNet - Detection network) [33]. Với kiến trúc là các phép lặp của thuật toán giảm độ dốc dự kiến kết (gradient descent) hợp thành một mạng. DetNet đã cho kết quả về độ chính xác vượt trội các phương pháp nhận dạng tuyến tính ở mức BER đạt 10^{-3} dB tại tỷ số tín trên tạp (SNR - Signal noise ratio) 10 dB. Tuy nhiên, do số lượng tham số cần đào tạo là lớn nên quá trình đào tạo có thể tốn chi phí, từ đó một số mạng ISD đã được đề xuất [18, 21], với độ chính xác tốt hơn DetNet nhưng số lượng tham số đào tạo chỉ dưới 100. Ngoài ra, rất nhiều các mô hình mạng NN khác đã được đề xuất, như mạng trí nhớ dài hạn/ngắn hạn (LSTM - Long/short-term memory), bộ tự mã hoá (Autoencoders), ... [42].

Tương tự như hai phương pháp kể trên, RL cũng được đưa sang ứng dụng cho nhận dạng kênh truyền. Trong nghiên cứu [25], nhóm các tác giả đã trình bày một phương pháp khử nhiễu trên miền tần số dựa trên RL không cần kiến thức kênh tiền nghiệm và dữ liệu được dán nhãn trước. Cụ thể, thuật toán cung cấp một cải tiến đáng kể so với phương pháp ước lượng bình phương nhỏ nhất thực tế (LS - Least squares) và mang lại hiệu suất tiệm cận với ước lượng lỗi bình phương trung bình nhỏ nhất tuyến tính (LMMSE - Linear MMSE) lý tưởng với toàn bộ thông tin về trạng thái kênh (CSI - Channel state information).

CHƯƠNG 2

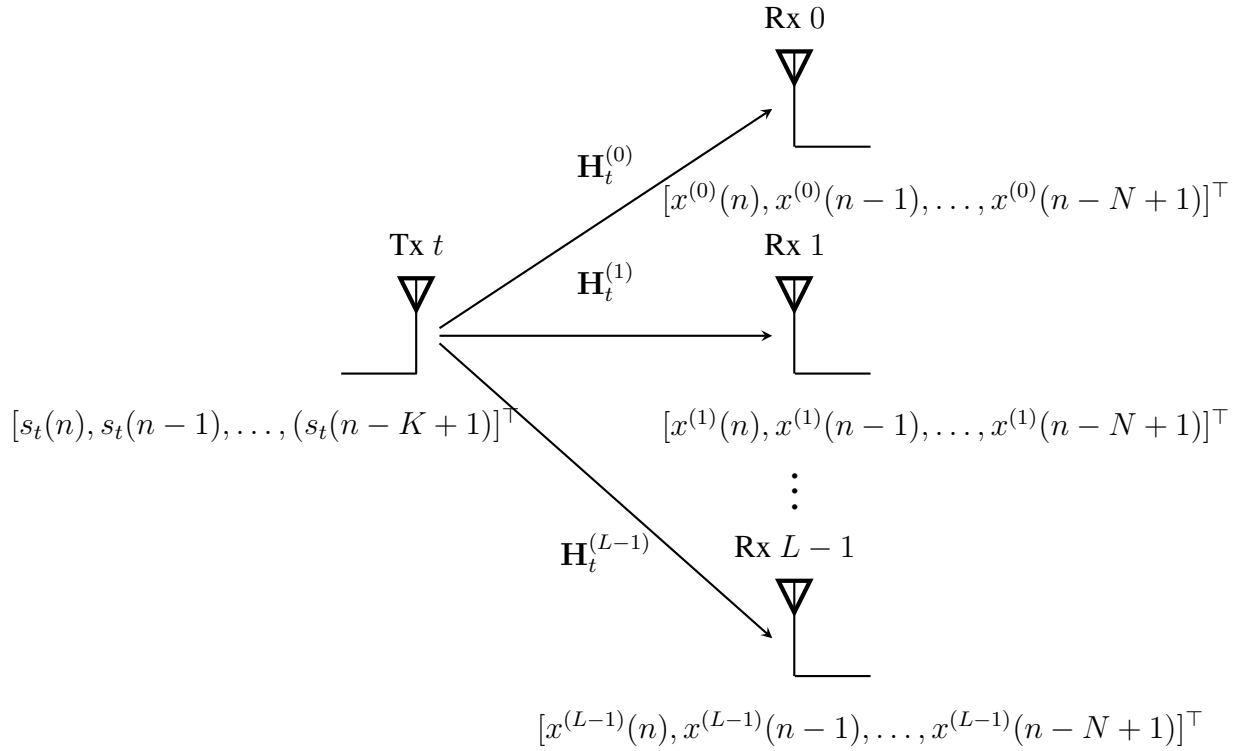
NHẬN DẠNG HỆ THỐNG SỬ DỤNG PHƯƠNG PHÁP BÁN MÙ

Trong chương này, trước hết, tác giả sẽ trình bày các tìm hiểu sơ lược về một thuật toán ước lượng kênh truyền mù truyền thống là bộ cân bằng kênh tham chiếu (MRE - Mutually referenced equalizers). Giải thuật B-MRE sẽ được phát triển để hoạt động với các hệ thống MIMO. Sau đó, phương pháp SB-MRE được đề xuất bằng cách sử dụng thêm một số lượng nhỏ pilot cùng với thông tin từ B-MRE. Sau đó, các hai phương pháp nhằm giảm thiểu chi phí của thuật toán SB-MRE được đề xuất thông qua việc giảm độ phức tạp của B-MRE. Cuối cùng, các mô phỏng được thực hiện để kiểm chứng thuật toán đề xuất và đưa ra kết luận.

2.1. Sơ lược về thuật toán B-MRE

Trong luận văn này, thuật toán bộ cân bằng kênh tham chiếu (MRE - Mutually referenced equalizers) [8] được lựa chọn để phát triển theo hướng tiếp cận SB. Giải thuật gốc đã được đề xuất từ năm 1997 bởi Gesbert D. và các đồng tác giả. Lý do lựa chọn MRE là phương pháp này có tốc độ xử lý nhanh, đảm bảo điều kiện hội tụ toàn phần, và đặc biệt có nhiều phương pháp thực thi, có thể kể đến như nhóm (batch), trung bình bình phương nhỏ nhất (LMS - Least mean squares), Kể từ năm 1997, nhiều công bố đã được đề xuất nhằm cải thiện giải thuật B-MRE. Trong [7], các tác giả đã đề xuất phương pháp thực thi bình phương nhỏ nhất đệ quy (RLS - Recursive least squares) hiệu quả cho B-MRE. Tiếp đến, Gasbert D. và các đồng tác giả [9] tiếp tục trình bày ý tưởng về một phiên bản cải tiến hỗ trợ MIMO của B-MRE. Năm 2000, J. van der Veen và các cộng sự [41] đã đề xuất cải thiện hiệu suất của phương pháp nhận dạng mù bằng việc kết hợp thuật toán B-MRE với một thuật toán mù khác là CMA [39] cho các hệ thống SIMO. Đến năm 2015, [47] đã trình bày ý tưởng về việc giảm thiểu số lượng bộ lọc cần ước lượng của B-MRE từ K về chỉ 2 bộ lọc. Qua đó giảm đáng kể độ phức tạp của thuật toán tuy nhiên phải đánh đổi một phần độ chính xác. Sau hơn 20 năm phát triển, B-MRE vẫn còn các điểm hạn chế cố hữu cần được giải quyết như sau

1. Yêu cầu khả năng tính toán lớn khi số lượng các bộ lọc cần ước lượng tăng, ví dụ số lượng các mẫu trong bộ đệm N lớn. Hơn nữa, đa số các phiên bản B-MRE được đề xuất kể trên chưa hỗ trợ MIMO.
2. Yêu cầu một số thông tin về kênh truyền ví dụ bậc của kênh và độ trễ giữa các ăng-



Hình 2.1. Minh hoạ hệ thống SIMO sử dụng phương pháp nhận dạng B-MRE.

ten thu. Hơn nữa, độ chính xác của B-MRE cũng chưa thể so sánh với các phương pháp NB.

Từ các phân tích kể trên, trong phần tiếp theo, thuật toán B-MRE gốc được sửa đổi để hỗ trợ MIMO sẽ được trình bày. Sau đó là phương pháp SB-MRE được đề xuất nhằm khắc phục các hạn chế của B-MRE kể trên.

Về cơ bản, MRE là phương pháp sử dụng các bộ làm bằng tuyến tính để khôi phục tín hiệu của bên phát tương tự như ZF và MMSE đã trình bày ở chương 1. Tuy nhiên, thay vì chỉ ước lượng một bộ lọc để làm bằng kênh, MRE sẽ tìm KT bộ lọc để làm bằng mỗi kênh truyền giữa ăng-ten phát thứ t và tương ứng với độ trễ i . Ký hiệu $\mathbf{g}_{t,i} \in \mathbb{C}^{LN \times 1}$ là bộ lọc tương ứng độ trễ i và ăng-ten phát thứ t . Với $i = 0, \dots, K-1$, tại thời điểm n , phương trình cân bằng kênh truyền sử dụng MRE như sau

$$\mathbf{g}_{t,i}^H * \mathbf{x}(n) = \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} g_{t,i}^H(k) x^{(l)}(n-k) \approx \mathbf{s}_t(n-i) \quad (2.1)$$

$$\mathbf{g}_{t,i} = [g_{t,i}^{(0)}(0), \dots, g_{t,i}^{(0)}(N-1), \dots, g_{t,i}^{(L-1)}(0), \dots, g_{t,i}^{(L-1)}(N-1)]^\top \quad (2.2)$$

Tổng hợp lại K độ trễ, ma trận cân bằng cho bộ phát thứ t là $\mathbf{G}_t \in \mathbb{C}^{LN \times K}$ và có dạng

$$\mathbf{G}_t = [\mathbf{g}_{t,0}, \dots, \mathbf{g}_{t,K-1}] \quad (2.3)$$

Trong trường hợp kênh truyền không có tạp âm AWGN, việc khôi phục các ký hiệu gửi đi có thể được thực hiện chính xác với ma trận cân bằng đầy đủ cho T bên phát, $\bar{\mathbf{G}}$, tương ứng là bất kỳ nghịch đảo chéo nào của \mathbf{H} . Với $\bar{\mathbf{G}}$, việc khôi phục tín hiệu được biểu diễn như sau

$$\begin{aligned} [\mathbf{G}_0, \dots, \mathbf{G}_{T-1}]^H \mathbf{x}(n) &= [\mathbf{s}_0^\top(n), \dots, \mathbf{s}_{T-1}^\top(n)]^\top \\ \bar{\mathbf{G}}^H \mathbf{x}(n) &= \bar{\mathbf{s}}(n) \end{aligned} \quad (2.4)$$

Trong trường hợp có sự xuất hiện của tạp âm AWGN, phương pháp MRE khai thác sự phân tập của độ trễ từ các kênh truyền khác nhau để tìm ra tất cả các nghịch đảo của kênh truyền. Trên hình 2.1 là minh hoạ của một hệ thống SIMO sử dụng phương pháp MRE. Tính chất phân tập độ trễ này được biểu diễn là

$$\mathbf{g}_i^H \mathbf{x}(n) = \mathbf{g}_{i+1}^H \mathbf{x}(n+1) \quad (2.5)$$

trong đó \mathbf{g} là dạng véc-tơ của ma trận làm bằng $\bar{\mathbf{G}}$ như được biểu diễn trong phương trình (2.13). Đầu ra của bộ làm bằng thứ i và $i+1$ “được tham chiếu” (referenced) lẫn nhau cho đến độ trễ chu kỳ của ký hiệu. Phương pháp MRE chọn hàm chi phí (cost function) của $\bar{\mathbf{G}}$ cho trường hợp không có ràng buộc như dưới đây

$$\mathcal{L}(\bar{\mathbf{G}}) = \mathbf{g}^H \mathbf{R} \mathbf{g} \quad (2.6)$$

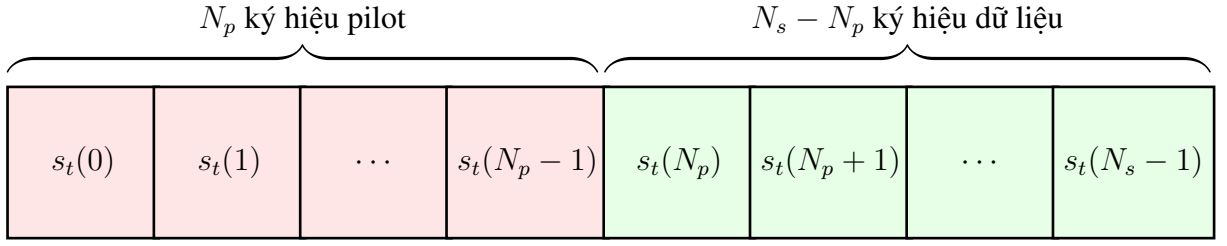
trong đó ma trận $\mathbf{R} \in \mathbb{C}^{LNKT \times LNKT}$ được tạo thành từ hai ma trận của tín hiệu thu được gồm $\mathbf{x}(n)$ và $\mathbf{x}(n+1)$. \mathbf{R} có thể được biểu diễn như sau

$$\mathbf{R} \stackrel{\text{def}}{=} E \left(\mathbf{U}^H \mathbf{U} \right) \quad (2.7)$$

trong đó E kỳ vọng thống kê, \mathbf{U} được biểu diễn như phương trình (2.8) với $\mathbf{I}_{T(K-1)}$ là các ma trận đơn vị có kích thước $T(K-1) \times T(K-1)$ và $\mathbf{0}$ là véc-tơ cột có kích thước $T \times 1$.

$$\mathbf{U} = (\mathbf{I}_{T(K-1)}, \mathbf{0}) \otimes \mathbf{x}^H(n) - (\mathbf{0}, \mathbf{I}_{T(K-1)}) \otimes \mathbf{x}^H(n+1) \quad (2.8)$$

Đến đây, tùy thuộc vào ràng buộc C1, C2, hoặc C3, việc ước lượng \mathbf{g} sẽ được thực hiện khác nhau. Cụ thể, C1 là ràng buộc bậc 2 (quadratic) đơn giản nhất, chỉ để đảm bảo tính chất đã được đưa ra ở phương trình (2.5). Tiếp đến, C2 là ràng buộc tuyến tính (linear) yêu cầu $\text{trace}(\mathbf{U}^H \bar{\mathbf{G}}) = 1$. Cuối cùng, C3 là ràng buộc giữ cho công suất đầu ra của bộ cân bằng không đổi, C3 yêu cầu $\mathbb{E} \|\bar{\mathbf{G}}^H \mathbf{x}(n)\|^2 = 1$. Trong luận văn này, ràng buộc C1 quadratic đơn giản nhất sẽ được lựa chọn. Qua đó, giá trị duy nhất nhỏ nhất và



Hình 2.2. Minh hoạ chuỗi một ký hiệu s_t sử dụng cho SB-MRE.

ổn định \mathbf{g} thu được bằng cách chọn giá trị véc-tơ riêng nhỏ nhất của ma trận \mathbf{R} .

$$\mathbf{g} = \mathbf{v}^\downarrow(\mathbf{R}) \quad (2.9)$$

2.2. Đề xuất phương pháp nhận dạng hệ thống SB-MRE cho MIMO

Phương pháp nhận dạng hệ thống SB-MRE được đề xuất dựa trên hướng tiếp cận đơn giản nhất của SB như đã trình bày ở mục 1.4 đó là sử dụng thêm một phần nhỏ pilot để kết hợp với B-MRE. Đi vào chi tiết, giả sử tại bộ phát thứ t , các chuỗi ký hiệu s_t có độ dài N_s sẽ được truyền đi. Trong đó, s_t bao gồm N_p ký hiệu pilot và $N_s - N_p$ ký hiệu dữ liệu như minh hoạ trên hình 2.2, được biểu diễn như sau

$$\mathbf{s}_t(n) = [s_t(n), \dots, s_t(n - N_p + 1), s_t(n - N_p), \dots, s_t(n - N_s + 1)]^\top \quad (2.10)$$

Các ký hiệu pilot sẽ được thu thập ở bên nhận, và sử dụng để nhận dạng kênh truyền theo hướng tiếp cận mù. Trong nghiên cứu này, phương pháp bình phương nhỏ nhất (LS - Least squares) được đề xuất để tìm ra ma trận nghịch đảo $\bar{\mathbf{G}}$ của \mathbf{H} . Giải thuật LS được biểu diễn như dưới đây

$$\hat{\mathbf{G}} = \arg \min_{\bar{\mathbf{G}} \in \mathbb{C}^{LN \times KT}} \sum_{i=N-1}^{N_p-1} \|\bar{\mathbf{s}}(n) - \bar{\mathbf{G}}^H \mathbf{x}(n)\|_F^2 \quad (2.11)$$

trong đó $\|\cdot\|_F$ là Frobenius norm. Như đã trình bày ở trên, hàm mất mát của $\bar{\mathbf{G}}$ cho giải thuật B-MRE được đưa ra ở phương trình (2.6). Để kết hợp thành phần pilot như trên phương trình (2.11) với thành phần B-MRE kể trên, một hàm mất mát mới kết hợp tự B-MRE và LS dựa trên pilot được đề xuất dựa trên phương pháp Lagrange multiplier [4]. Hàm mất mát chung của SB-MRE sẽ là

$$\mathcal{L}(\bar{\mathbf{G}}) = \sum_{i=N-1}^{N_p-1} \|\bar{\mathbf{s}}(n) - \bar{\mathbf{G}}^H \mathbf{x}(n)\|_F^2 + \lambda \mathbf{g}^H \mathbf{R} \mathbf{g} \quad (2.12)$$

trong đó λ là trọng số của thành phần mù trong SB-MRE, \mathbf{R} của B-MRE được đưa ra ở

công thức (2.7) cho ràng buộc bậc hai. \mathbf{g} là dạng véc-tơ của ma trận làm bằng kênh $\bar{\mathbf{G}}$ được biểu diễn như sau

$$\begin{aligned}\mathbf{g} &= \text{vec}(\bar{\mathbf{G}}) = \left[\bar{\mathbf{G}}_0^\top, \bar{\mathbf{G}}_1^\top, \dots, \bar{\mathbf{G}}_{K-1}^\top \right]^\top \\ \bar{\mathbf{G}}_i &= \left[\mathbf{g}_{0,i}^\top, \mathbf{g}_{1,i}^\top, \dots, \mathbf{g}_{T-1,i}^\top \right]^\top\end{aligned}\quad (2.13)$$

Không làm mất đi tính tổng quát, thành phần LS trong công thức (2.12) được chuyển vị và lấy liên hợp phức, toán tử \sum chuyển sang dạng ma trận của các tín hiệu nguồn và thu lần lượt là $\tilde{\mathbf{S}}$ và $\tilde{\mathbf{X}}$. Từ đó, hàm mất mát của $\bar{\mathbf{G}}$ sẽ có dạng là

$$\begin{aligned}\mathcal{L}(\bar{\mathbf{G}}) &= \sum_{i=N-1}^{N_p-1} \left\| \bar{\mathbf{s}}^H(n) - \mathbf{x}^H(n) \bar{\mathbf{G}} \right\|_F^2 + \lambda \mathbf{g}^H \mathbf{R} \mathbf{g} \\ &= \left\| \tilde{\mathbf{S}}^H - \tilde{\mathbf{X}}^H \bar{\mathbf{G}} \right\|_F^2 + \lambda \mathbf{g}^H \mathbf{R} \mathbf{g}\end{aligned}\quad (2.14)$$

trong đó $\tilde{\mathbf{S}}, \tilde{\mathbf{X}}$ là các ma trận có kích thước lần lượt $\mathbb{C}^{KT \times (N_p - N + 1)}$ và $\mathbb{C}^{LN \times (N_p - N + 1)}$. Các ma trận này có dạng như bên dưới

$$\begin{aligned}\tilde{\mathbf{S}} &= [\bar{\mathbf{s}}(N-1), \dots, \bar{\mathbf{s}}(N_p-1)] \\ \tilde{\mathbf{X}} &= [\mathbf{x}(N-1), \dots, \mathbf{x}(N_p-1)]\end{aligned}\quad (2.15)$$

Sử dụng thuộc tính của toán tử véc [23] với ba ma trận $\mathbf{A}, \mathbf{B}, \mathbf{X}$ như sau

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}^\top \otimes \mathbf{A}) * \text{vec}(\mathbf{X}) \quad (2.16)$$

biểu diễn LS được véc-tơ hoá, từ đó, dạng của hàm mất mát SB-MRE là

$$\begin{aligned}\mathcal{L}(\mathbf{g}) &= \left\| \text{vec}(\tilde{\mathbf{S}}^H) - (\mathbf{I}_{KT} \otimes \tilde{\mathbf{X}}^H) \text{vec}(\bar{\mathbf{G}}) \right\|_F^2 + \lambda \mathbf{g}^H \mathbf{R} \mathbf{g} \\ &= \left\| \bar{\mathbf{s}} - \mathbf{A} \mathbf{g} \right\|_F^2 + \lambda \mathbf{g}^H \mathbf{R} \mathbf{g} \\ &= \mathbf{g}^H \mathbf{A}^H \mathbf{A} \mathbf{g} + \left\| \bar{\mathbf{s}} \right\|_F^2 - 2 \Re(\mathbf{g}^H \mathbf{A}^H \bar{\mathbf{s}}) + \lambda \mathbf{g}^H \mathbf{R} \mathbf{g}\end{aligned}\quad (2.17)$$

Để tìm giá trị cực tiểu của hàm mất mát trên phương trình (2.17), lấy đạo hàm riêng của $\mathcal{L}(\mathbf{g})$ theo \mathbf{g} như sau

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{g}}(\mathbf{g}) &= 0 \\ (\mathbf{A}^H \mathbf{A} + \lambda \mathbf{R}) \mathbf{g} &= \mathbf{A}^H \bar{\mathbf{s}}\end{aligned}\quad (2.18)$$

Giá trị cuối cùng của ma trận cân bằng kênh $\bar{\mathbf{G}}$ dưới dạng véc-tơ \mathbf{g}_{SB} sử dụng

phương pháp SB-MRE thu được là

$$\mathbf{g}_{SB} = \left(\mathbf{A}^H \mathbf{A} + \lambda \mathbf{R} \right)^{-1} \mathbf{A}^H \bar{\mathbf{s}} \quad (2.19)$$

2.3. Đề xuất giảm thiểu chi phí của thuật toán SB-MRE

Trong đề xuất gốc và đề xuất SB-MRE đã trình bày ở trên, độ phức tạp của thành phần B-MRE là $\mathcal{O}(LNKT)$ [8]. Trong đó KT bộ cân bằng kênh sẽ được ước lượng, tuy nhiên chỉ một được sử dụng sau đó. Điều này làm ra tăng lượng tài nguyên tính toán không cần thiết đặc biệt là khi N lớn. Do đó, trong nghiên cứu này, số lượng các bộ lọc phải ước lượng sẽ được giảm từ K còn 2 bộ lọc, bao gồm bộ lọc không trễ (thứ tự 0) và bộ lọc có độ trễ lớn nhất (thứ tự $K - 1$). Giải thuật B-MRE sau khi đã được giảm độ phức tạp sẽ được viết ngắn gọn là ‘B-MRE rc’ (reduced cost) và giải thuật SB-MRE sử dụng thông tin từ B-MRE rc sẽ được viết tắt là ‘SB-MRE rc’. Với đề xuất này, độ phức tạp của B-MRE được giảm về $\mathcal{O}(LNT)$, và ma trận cân bằng kênh truyền cho bộ phát thứ t sẽ có dạng

$$\mathbf{V}_t = [\mathbf{g}_{t,0}, \mathbf{g}_{t,K-1}] \quad (2.20)$$

theo đó, tín hiệu nguồn được ước lượng cho ăng-ten phát thứ t được biểu diễn như sau

$$\mathbf{V}_t^H \mathbf{x}(n) = [s_t(n), s_t(n - K + 1)]^T = \mathbf{s}_t(n) \quad (2.21)$$

cuối cùng, việc tính toán toàn bộ bậc của ma trận \mathbf{R} là không cần thiết. Phương trình (2.7) được sửa đổi trở thành

$$\mathbf{U} = (\mathbf{I}_T, \mathbf{0}) \otimes \mathbf{x}^H(n) - (\mathbf{0}, \mathbf{I}_T) \otimes \mathbf{x}^H(n + K - 1) \quad (2.22)$$

2.4. Mô phỏng và đánh giá

Bảng 2.1. Các tham số mô phỏng hệ thống truyền thông không dây của SB-MRE.

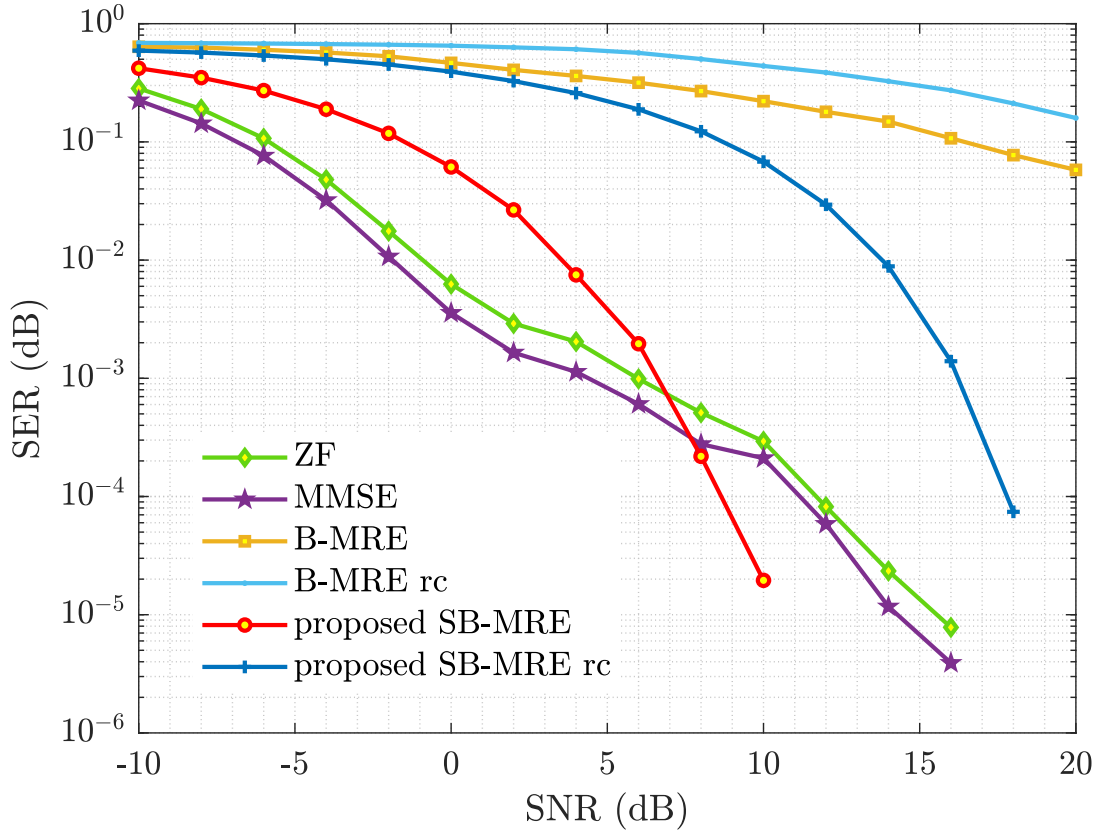
Thông số mô phỏng	Giá trị
Kích thước hệ thống MIMO	$T = 2, L = 4$
Loại điều chế	QPSK (Quadrature phase shift keying)
Bậc của kênh truyền	$M = 3$
Kích thước mỗi cửa sổ mẫu	$N = 10$
Kích thước của một chuỗi ký hiệu	$N_s = 256$
Số lượng ký hiệu pilot	$N_p = 32$
Số lượng bộ làm bằng của B-MRE rc	2
Trọng số của thành phần B-MRE	$\lambda = 0,1$

Để kiểm chứng hiệu suất hoạt động của thuật toán SB-MRE được đề xuất, các tham số mô phỏng trong bảng 2.1 được áp dụng cho kênh truyền vô tuyến. Các mô phỏng chạy trên Matlab, và lấy giá trị trung bình của 1.000 lần chạy làm kết quả cuối cùng. Thông số đánh giá độ chính xác của việc nhận dạng là tỷ lệ sai số ký hiệu (SER - Symbol error rate)

$$\text{SER} = \frac{1}{1000} \sum_{K=1}^{1000} \frac{N_e}{N_s} \quad (2.23)$$

với K là thứ tự chạy mô phỏng và N_e là số lượng các ký tự ước lượng sai. Đầu tiên, độ chính xác của SB-MRE được so sánh với các thuật toán ước lượng kênh truyền tuyến tính bao gồm ZF và MMSE đã trình bày ở chương 1. Trên hình 2.3, kết quả mô phỏng cho thấy độ chính xác của SB-MRE, SB-MRE rc, ZF, và MMSE là vượt trội khi so sánh với thuật toán mù truyền thống B-MRE và B-MRE rc. Ở các mức SNR thấp dưới 5 dB, SER của ZF và MMSE là vượt trội khi so sánh với B-MRE, SB-MRE ở dạng đầy đủ độ phức tạp của B-MRE hay đã giảm bớt chỉ còn 2 bộ làm bằng. Nguyên nhân là do, ở các mức SNR thấp, hiệu suất của thành phần B-MRE là không đáng kể. Tại giá trị SNR = 7 dB, phương pháp SB-MRE đã chạm đến hiệu suất ngang bằng với ZF và MMSE ở mức gần đến SER $\approx 10^{-4}$ dB và vượt trội hoàn toàn so với B-MRE truyền thống. Khi SNR tiếp tục tăng, tỷ lệ chính xác của SB-MRE sẽ vượt trội so với ZF và MMSE. Tuy nhiên, lưu ý rằng, SB-MRE chỉ cần 32/256 ký hiệu pilot để đạt được hiệu suất kể trên, trong khi ZF và MMSE cần đến thông tin trạng thái kênh truyền (H) chính xác để được kết quả như trong mô phỏng. Khi giảm đi số lượng bộ lọc của B-MRE, SER của phương pháp SB-MRE rc cũng ở mức có thể chấp nhận được khi SER giảm dần khi SNR tăng, tuy nhiên SER thu được vẫn là thấp hơn SB-MRE, ZF hay MMSE. Hiệu suất của B-MRE dù ở dạng đầy đủ K bộ làm bằng hay giảm xuống chỉ 2 bộ lọc đều ở mức thấp, chỉ đạt SER $\approx 10^{-1}$ dB ở mức SNR = 20 dB. Xét về mức SNR cần để các thuật toán đạt được SER tuyệt đối, hay việc ước lượng trong mô phỏng là không có lỗi, phương pháp đề xuất SB-MRE cần 11 dB, SB-MRE rc tại 19 dB, ZF và MMSE đạt được tại 17 dB, ngược lại B-MRE và B-MRE rc không đạt được mức SER tuyệt đối. Qua đó, có thể nhận xét giải thuật SB-MRE đã cho kết quả như đã đề ra khi vừa giảm thiểu được số lượng pilot cần dùng mà vẫn bù lại được độ chính xác từ thành phần B-MRE. Việc giảm số lượng bộ lọc của B-MRE vẫn sẽ cho kết quả SB-MRE rc chấp nhận được khi SNR ở mức cao.

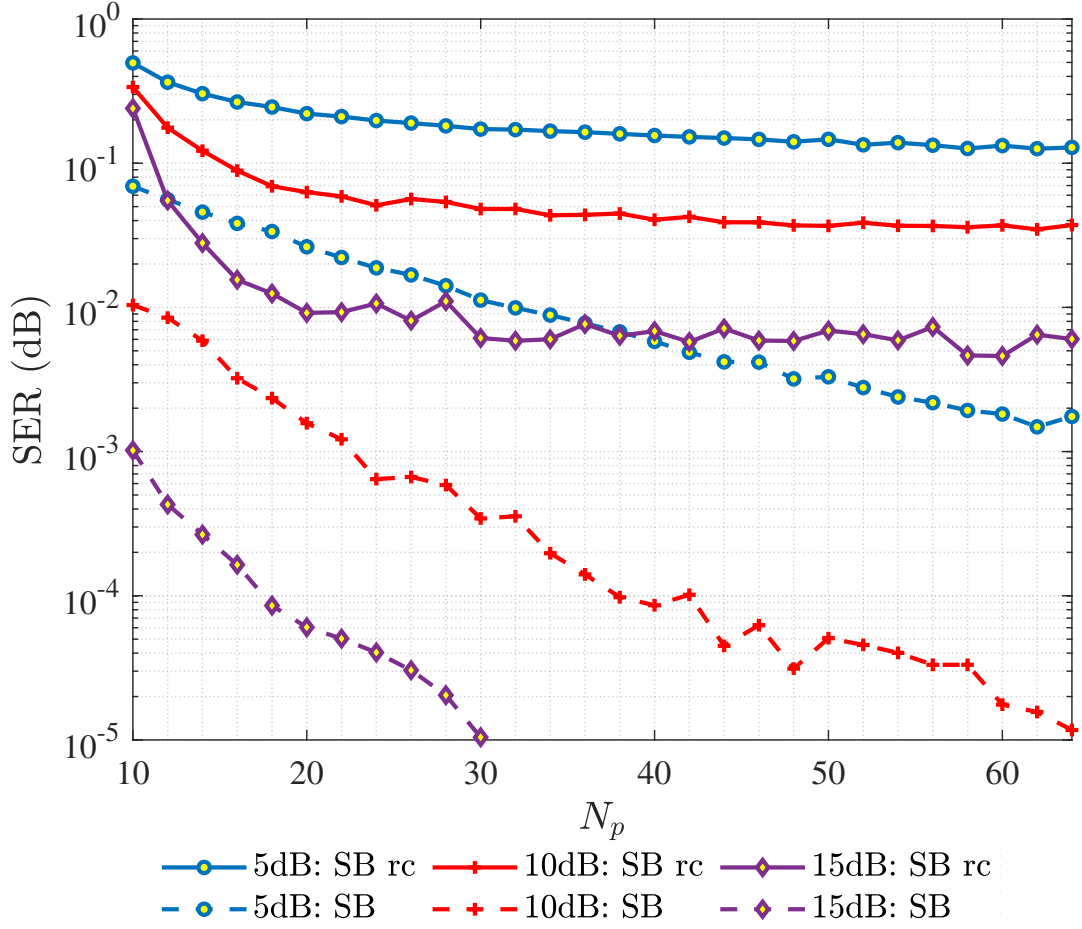
Tiếp đến, để kiểm chứng sự ảnh hưởng của số lượng ký hiệu pilot (N_p) đến hiệu suất của phương pháp đề xuất SB-MRE và SB-MRE rc, SER tương ứng với các N_p được mô phỏng và đưa ra trong hình 2.4. Số lượng pilot được thay đổi từ $N \rightarrow N_s/4$ tương ứng là $10 \rightarrow 64$ và SNR được cố định là 5, 10 hoặc 15 dB. Dễ nhận thấy, khi số lượng ký hiệu



Hình 2.3. So sánh SER của thuật toán đề xuất SB-MRE với các phương pháp nhận dạng tuyến tính.

pilot và SNR tăng thì SER sẽ giảm, tuy nhiên độ suy giảm là khác nhau. Với SB-MRE rc, tỷ lệ lỗi ký hiệu sẽ giảm nhanh khi N_p tăng từ 10 \rightarrow 20 ký hiệu, và giảm chậm và gần như hội tụ khi N_p tiếp tục tăng lên đến 64. Các mức SNR lớn hơn sẽ cho SER tốt hơn và có thể đạt thấp hơn 10^{-2} cho SB-MRE giảm chi phí với SNR = 15 dB, $N_p > 20$ ký hiệu. Với SB-MRE, trước hết hiệu suất của SB-MRE sẽ vượt trội trên 10^1 dB so với SB-MRE đã giảm chi phí ở tất cả các giá trị SNR khác nhau. Thứ hai, nhìn chung SER sẽ giảm dần đều ở tất cả các giá trị N_p . Thứ ba, độ dốc của các đường SER là tăng dần với SNR lớn hơn. Ở giá trị SNR = 15 dB, chỉ cần 31 ký hiệu pilot để SB-MRE đạt được SER tuyệt đối. Tóm lại, số lượng pilot ảnh hưởng lớn đến độ chính xác của SB-MRE và ít hơn cho SB-MRE rc.

Cuối cùng, ảnh hưởng của thành phần B-MRE trong SB-MRE sẽ được khảo sát thông qua trọng số λ . Trên hình 2.5, hai phương pháp được đề xuất SB-MRE và SB-MRE rc sẽ được mô phỏng với các λ và SNR khác nhau. Cụ thể, giá trị λ sẽ được thay đổi trong khoảng 0,01 \rightarrow 0,2 với bước 0,01 và SNR được cố định là 5, 10 hoặc 15 dB. Có thể thấy rằng, SER của SB-MRE và SB-MRE rc không bị ảnh hưởng bởi λ nhiều như số lượng pilot. Đặc biệt khi SNR thấp như 5 dB, SER giảm rất ít khi λ tăng, do ảnh hưởng

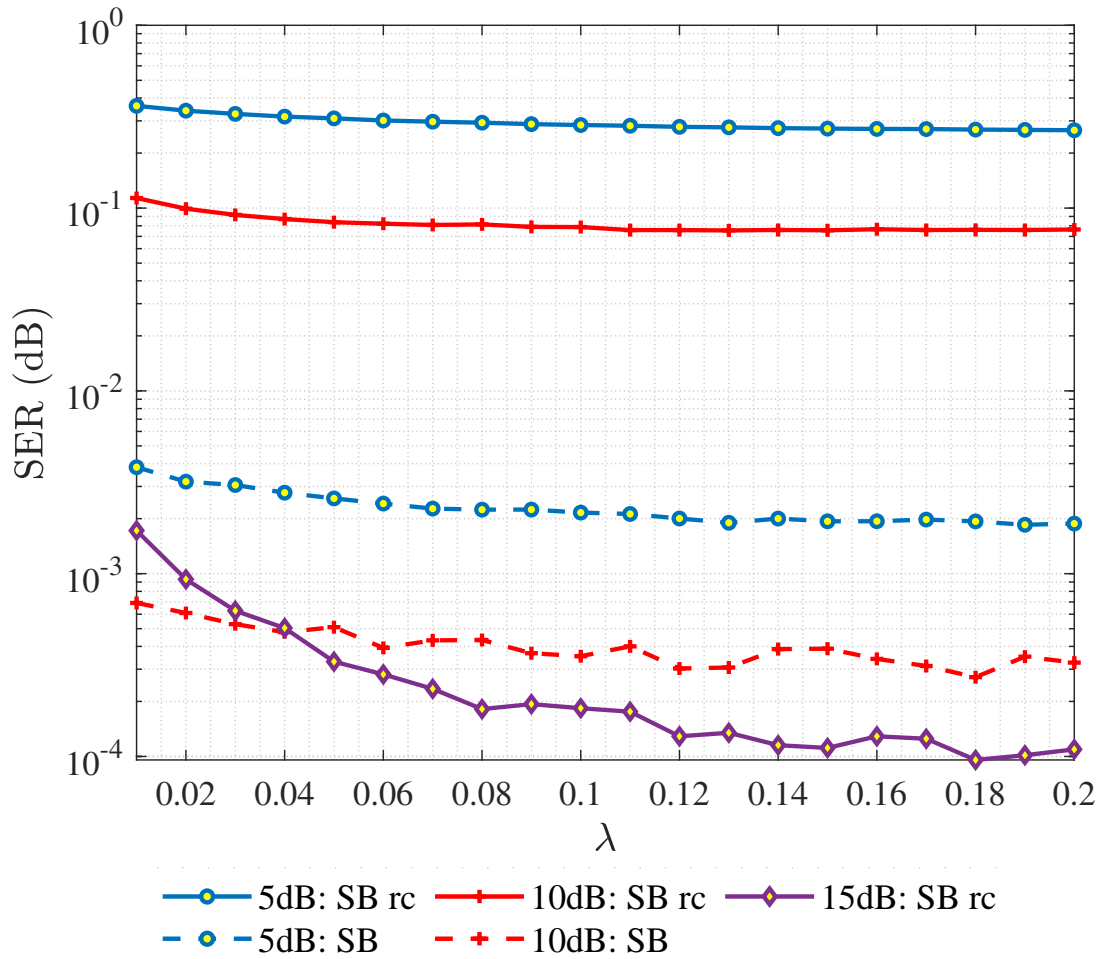


Hình 2.4. So sánh SER của thuật toán đề xuất SB-MRE với các số lượng pilot (N_p) và SNR khác nhau.

suất của B-MRE ở các mức SNR thấp như đã khảo sát trên hình 2.3. Khi SNR cao hơn như 10 dB, B-MRE bắt đầu ảnh hưởng đến độ chính xác chung, đặc biệt là với SB-MRE khi ước lượng toàn bộ K bộ làm bằng khi độ dốc của đường SER là rõ ràng hơn. Tuy nhiên, phải đến khi SNR = 15 dB, B-MRE mới thực sự gây ra ảnh hưởng, đường SER của SB-MRE rc giảm rõ rệt từ $10^{-3} \rightarrow 10^{-4}$ khi λ tăng lên. Ở mức SNR này, SB-MRE đã đạt độ chính xác tuyệt đối chỉ với thành phần pilot $N_p = 32$ như đã khảo sát trên hình 2.3 và 2.4 nên không được biểu diễn trên hình 2.5. Qua đó, có thể nhận xét trọng số của thành phần mù có ảnh hưởng đến độ chính xác của SB-MRE nhưng chỉ khi tỷ lệ tín hiệu tạp âm ở mức cao để thành phần mù bắt đầu cho thấy tác dụng.

2.5. Kết luận chương

Trong chương này, một kỹ thuật nhận dạng kênh truyền sử dụng thuật toán SB-MRE do chúng tôi đề xuất đã được trình bày. Trước hết, tác giả đã trình bày ngắn gọn lại phương pháp B-MRE truyền thống sử dụng cho các hệ thống truyền thông MIMO. Sau đó, phương pháp SB-MRE được đề xuất bằng các kết hợp thông tin từ một số kỹ thuật



Hình 2.5. So sánh SER của thuật toán đề xuất SB-MRE với trọng số thành phần B-MRE (λ) và SNR khác nhau.

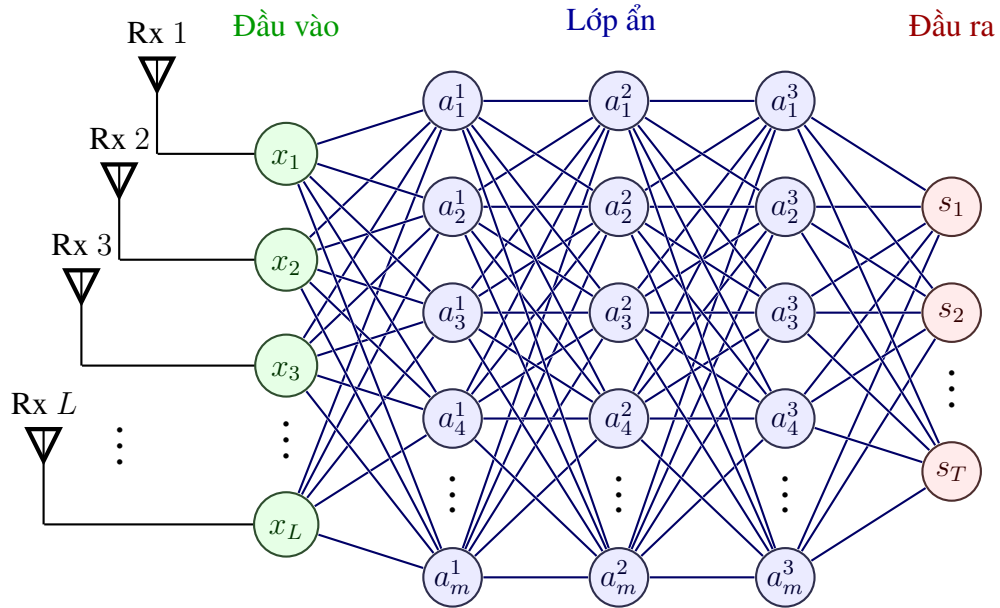
pilot trong chuỗi ký hiệu và thông tin từ thành phần B-MRE kể trên. Các phương pháp giảm thiểu chi phí của thuật toán SB-MRE cũng được làm rõ, bao gồm, giảm thiểu độ phức tạp của thành phần B-MRE thông qua việc giảm số lượng bộ lọc cần ước lượng và giảm số lượng số ký hiệu pilot cần sử dụng. Các kết quả mô phỏng đã chỉ ra hiệu suất của phương pháp SB-MRE được đề xuất là khá vượt trội khi so sánh với các thuật toán nhận dạng hệ thống tuyến tính cổ điển và B-MRE gốc. Các mô phỏng cũng được thực hiện để xác nhận sự ảnh hưởng của số lượng ký hiệu pilot và đóng góp của thành phần mù đến độ chính xác của thuật toán SB-MRE ở các mức SNR khác nhau. Từ đó, có thể thấy phương pháp SB-MRE là tiềm năng cho các hệ thống truyền thông MIMO/mMIMO khi vừa giảm được số lượng ký hiệu pilot cần cho việc nhận dạng hệ thống mà vẫn giữ được một độ chính xác chấp nhận được.

CHƯƠNG 3

NHẬN DẠNG HỆ THỐNG SỬ DỤNG MẠNG HỌC SÂU

Trong chương này, tác giả sẽ đề xuất một mô hình mạng học sâu ISDNN sử dụng cho nhận dạng hệ thống viễn thông MIMO/mMIMO. Đầu tiên, sơ lược về hai hướng tiếp cận sử dụng mạng nơ-ron sâu sẽ được giới thiệu. Kế đến là khái niệm về kỹ thuật mở rộng sâu (Deep unfolding). Một mô hình được mở rộng sâu từ bộ tối ưu hợp lệ cực đại (MLE) là DetNet được trình bày để so sánh ở mục các kết quả mô phỏng. Tiếp đến, từ một giải thuật ISD đã được đề xuất trong [21], kết hợp với cách tiếp cận mở rộng sâu tại [18], tác giả đề xuất một mạng nơ-ron sâu ISDNN để nhận dạng hệ thống. Các bước mô phỏng và đánh giá sẽ được đưa ra để cho thấy tiềm năng của phương pháp đề xuất và kết luận của chương.

3.1. Giới thiệu về mạng nơ-ron sâu và mở rộng sâu (Deep unfolding)



Hình 3.1. Minh họa sử dụng DNN để nhận dạng hệ thống viễn thông.

Trong chương 1, các phương pháp nhận dạng hệ thống sử dụng các phương pháp ML/DL đã được chia làm ba loại, trong đó phương pháp sử dụng các mạng nơ-ron đang được quan tâm nghiên cứu. Các mạng nơ-ron sâu (DNN - Deep-neural network) đã được sử dụng rộng rãi trong các ứng dụng như xử lý tiếng nói, ngôn ngữ tự nhiên, hình ảnh, thị giác máy, trò chơi trực tuyến [31]. Mười năm trở lại đây, đã có nhiều nghiên cứu ứng dụng các mô hình mạng DNN khác nhau cho vấn đề nhận dạng hệ thống viễn thông

không đây. Trên hình 3.1 là một mô hình minh họa việc sử dụng DNN để ước lượng kênh truyền và khôi phục tín hiệu gốc. Có thể chia các phương pháp này thành hai hướng tiếp cận, bao gồm hướng dữ liệu (data-driven) và hướng mô hình (model-driven) [18]. Các phương pháp data-driven trực tiếp học các đặc trưng từ một tập lớn các dữ liệu (dataset) để phục vụ cho các mục đích như ước lượng kênh truyền, phản hồi CSI, Tuy các phương pháp data-driven được đề xuất đều cho độ chính xác cao nhưng vẫn có những thách thức khi yêu cầu số lượng mẫu rất lớn và kéo theo đó là thời gian/chi phí cho việc đào tạo lớn. Các phương pháp model-driven [12] có thể một phần khắc phục các hạn chế này bằng việc tối ưu/đưa thêm các tham số học vào các mô hình có sẵn để kết hợp ưu điểm của data-driven và các mô hình toán học truyền thống.

Trong các năm gần đây, kỹ thuật mở rộng sâu (Deep unfolding) [45] là một giải pháp tiềm năng để chuyển các giải thuật truyền thống thành các kiến trúc mạng DNN theo hướng tiếp cận model-driven. Chi tiết về mở rộng sâu tại [13], các phương pháp yêu cầu các vòng lặp đi lặp lại (iterative inference) có thể dễ dàng chuyển đổi sang từng lớp của một mạng NN. Sau đó, sử dụng các giải thuật giảm dần độ dốc (GD - gradient descent) để đào tạo tham số trên các lớp mạng. Sau K lớp đào tạo tương tự như K vòng lặp trong thuật toán gốc, mô hình có thể đạt được mục tiêu mong muốn. Ví dụ, DetNet [32] là một mạng DNN dựa trên việc mở rộng sâu bộ nhận dạng MLE và sử dụng giảm dần độ dốc dự kiến (PGD - projected gradient descent) [5]. Trong mục tiếp theo, mô hình mạng nơ-ron sâu DetNet sẽ được giới thiệu ngắn gọn và kết quả của DetNet sẽ so sánh với mạng sẽ được đề xuất.

3.2. Mạng nơ-ron học sâu DetNet

Xét hệ thống MIMO/mMIMO tương tự đã trình bày ở hình 1.2. Tuy nhiên, thay vì mô hình hoá tín hiệu kênh truyền vô tuyến dưới dạng các bộ lọc FIR có chiều dài $M + 1$, trong DNN, giả sử: (i) T có thể coi là số lượng ăng-ten bên phát hoặc số lượng người dùng (user) với mỗi người dùng chỉ có một ăng-ten phát, (ii) ma trận \mathbf{H} là biến đổi tuyến tính của tín hiệu truyền thành tín hiệu nhận được tức $M = 0$, (iii) $N = 1$ tức mỗi ăng-ten nhận chỉ thu thập một ký hiệu tại một thời điểm, (iv) các ma trận sẽ được chuyển đổi sang dạng phần thực, ảo riêng biệt như phương trình (1.5) và (1.6). Biểu diễn đơn giản cho hệ thống MIMO/mMIMO như sau

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w} \quad (3.1)$$

trong đó, $\mathbf{H} \in \mathbb{R}^{2L \times 2T}$, $\mathbf{s} \in \mathbb{R}^{2T \times 1}$, $\mathbf{w} \in \mathbb{R}^{2L \times 1}$, và $\mathbf{x} \in \mathbb{R}^{2L \times 1}$. Để tìm bộ nhận dạng cho hệ thống kể trên, định nghĩa hàm mất mát $\mathcal{L}(\mathbf{s}; \hat{\mathbf{s}}_{\theta}(\mathbf{H}, \mathbf{x}))$ là khoảng cách giữa ký hiệu

gốc và ký hiệu được ước lượng. Tìm giá trị θ bằng cách tối thiểu hoá hàm mất mát kể trên.

$$\min_{\theta} \mathbb{E} \{ \mathcal{L}(\mathbf{s}; \hat{\mathbf{s}}_{\theta}(\mathbf{H}, \mathbf{x})) \} \quad (3.2)$$

Giải thuật tối ưu để giải quyết (3.2) là bộ ước lượng hợp lý cực đại (MLE - Maximum likelihood estimator) như sau

$$\hat{\mathbf{s}}_{\theta}(\mathbf{x}, \mathbf{H}) = \arg \min_{\mathbf{s} \in \mathbb{R}^{2T}} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2 \quad (3.3)$$

Tuy nhiên, độ phức tạp của MLE sẽ tăng theo cấp số mũ $\mathcal{O}(2^T)$ nên khó để triển khai trong các hệ mMIMO. Do vậy, DetNet được đề xuất nhằm tạo ra một kiến trúc mạng DNN đạt được tiệm cận độ chính xác với MLE. Trong nghiên cứu gốc, thay vì tạo ra một mạng nơ-ron nhằm ánh xạ trực tiếp từ \mathbf{x} về \mathbf{s} , việc phân tách \mathbf{x} thành các thành phần \mathbf{H} , \mathbf{s} , và \mathbf{w} sẽ cho hiệu quả cao hơn.

$$\mathbf{H}^{\top} \mathbf{x} = \mathbf{H}^{\top} \mathbf{H} \mathbf{s} + \mathbf{H}^{\top} \mathbf{w} \quad (3.4)$$

Kiến trúc DetNet dựa trên phương pháp PGD [5] cho việc tối ưu MLE như trên (3.3). Đạo hàm riêng được tách như trên (3.5) sử dụng luật chuỗi (chain rule) [23].

$$\begin{aligned} \hat{\mathbf{s}}_{k+1} &= \Pi \left[\hat{\mathbf{s}}_k - \delta_k \frac{\partial \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2}{\partial \mathbf{s}} \Big|_{\mathbf{s}=\hat{\mathbf{s}}_k} \right] \\ &= \Pi \left[\hat{\mathbf{s}}_k - \delta_k \mathbf{H}^{\top} \mathbf{x} + \delta_k \mathbf{H}^{\top} \mathbf{H} \hat{\mathbf{s}}_k \right] \end{aligned} \quad (3.5)$$

với \mathbf{s}_k là giá trị ước lượng tại lớp thứ k , $\Pi[\cdot]$ là một phép biến đổi phi tuyến tính, và δ_k là độ dài bước (step size) của quá trình học. Kiến trúc của mạng DetNet đề xuất trong [32] được biểu diễn như trên hình 3.2 và cách biểu diễn dưới dạng ma trận như sau

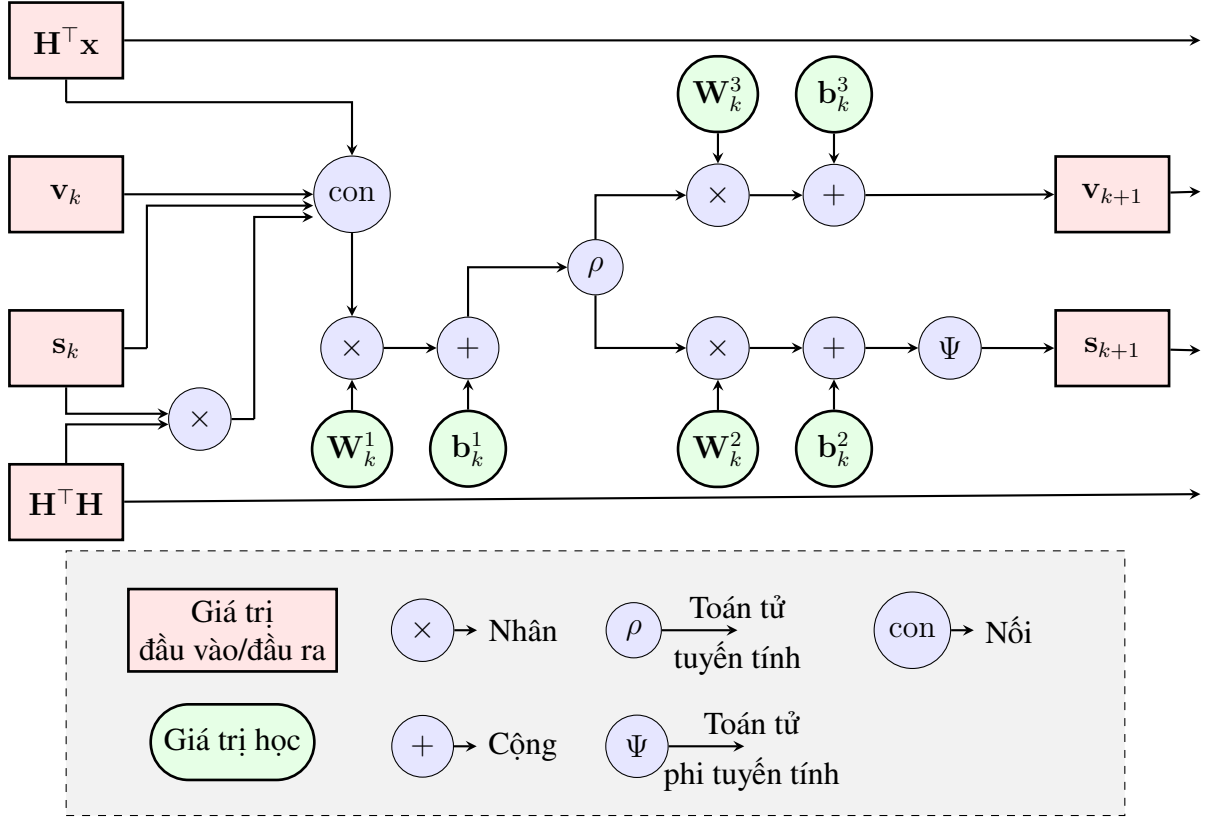
$$\mathbf{z}_k = \rho \left(\mathbf{W}_k^1 \begin{bmatrix} \mathbf{H}^{\top} \mathbf{x} \\ \hat{\mathbf{s}}_k \\ \mathbf{H}^{\top} \mathbf{H} \hat{\mathbf{s}}_k \\ \mathbf{v}_k \end{bmatrix} + \mathbf{b}_k^1 \right) \quad (3.6a)$$

$$\hat{\mathbf{s}}_{k+1} = \psi_{t_k} (\mathbf{W}_k^2 \mathbf{z}_k + \mathbf{b}_k^2) \quad (3.6b)$$

$$\hat{\mathbf{v}}_{k+1} = \mathbf{W}_k^3 \mathbf{z}_k + \mathbf{b}_k^3 \quad (3.6c)$$

$$\hat{\mathbf{s}}_1 = \mathbf{0} \quad (3.6d)$$

trong đó, $k = 1, \dots, K$ là số các lớp của mạng DetNet, ρ là một toán tử tuyến tính. ψ_{t_k}



Hình 3.2. Kiến trúc của một lớp trong mô hình mạng DetNet [32].

ký hiệu cho phép biến đổi phi tuyến tính phân đoạn, ở các mức t khác nhau, $\psi_{t_k}(s)$ được minh hoạ trên hình 3.3 và có biểu diễn toán học như sau

$$\psi_{t_k}(s) = -1 + \frac{\rho(s + t_k)}{|t_k|} - \frac{\rho(s - t_k)}{|t_k|} \quad (3.7)$$

Các tham số của việc học sẽ bao gồm

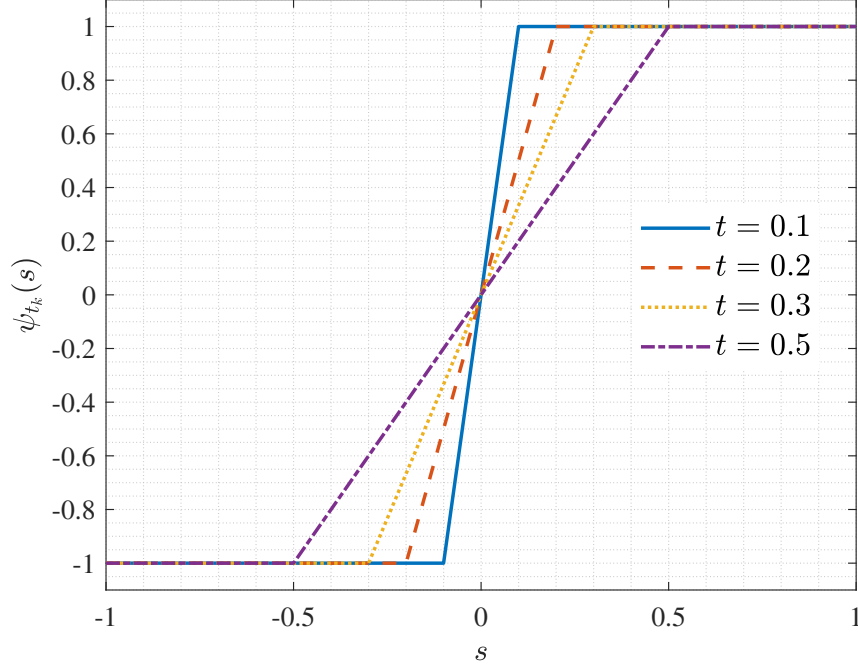
$$\theta = \{\mathbf{W}_k^1, \mathbf{b}_k^1, \mathbf{W}_k^2, \mathbf{b}_k^2, \mathbf{W}_k^3, \mathbf{b}_k^3, \mathbf{t}_k\}_{k=1}^K \quad (3.8)$$

Một hàm mất mát có dạng sai số toàn phương trung bình (MSE - Mean Squared Error) sẽ tổng hợp sai số từ kết quả đầu ra của tất cả các lớp để ước lượng sự hội tụ của mạng DetNet. Hàm mất mát như dưới đây

$$\mathcal{L}(\mathbf{s}; \hat{\mathbf{s}}(\mathbf{H}, \mathbf{x})) = \frac{1}{2T} \sum_{t=1}^{2T} \|s_t - \hat{s}_t\|^2 \quad (3.9)$$

3.3. Đề xuất mạng nơ-ron sâu ISDNN cho nhận dạng kênh truyền

Trong phần này, giải thuật của bộ nhận dạng ISD công bố tại [21] sẽ được trình bày. Từ đó, một mạng nơ-ron sâu ISDNN được đề xuất dựa trên kỹ thuật mở rộng sâu



Hình 3.3. Hàm phi tuyến tính phân đoạn $\psi_{t_k}(s)$ được sử dụng trong DetNet.

cho giải thuật ISD trước đó.

3.3.1. Bộ nhận dạng ISD cho hệ thống mMIMO

Giải thuật gốc tại [21] đã đề xuất một bộ nhận dạng kênh truyền tuần tự lặp lại gọi tắt là ISD để đạt được hiệu suất của MMSE với độ phức tạp thấp. Trong đó, bộ nhận dạng MMSE đã được chứng minh [30] có thể đạt được độ chính xác tiệm cận của MLE cho kênh đường lên cho các hệ mMIMO với $L/T \geq 10$. Lưu ý rằng, do \mathbf{H} là ma trận của các số thực nên phép chuyển vị liên hợp phức $(\cdot)^H$ sẽ được chuyển thành chuyển vị $(\cdot)^\top$.

$$\hat{\mathbf{s}}_{MMSE} = \left(\mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I}_{2T} \right)^{-1} \mathbf{H}^\top \mathbf{x} = \mathbf{P}^{-1} \mathbf{q} \quad (3.10)$$

ký hiệu $\mathbf{G}_\mathbf{H} = \mathbf{H}^\top \mathbf{H}$, $\mathbf{P} = \mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I}_{2T}$, và $\mathbf{q} = \mathbf{H}^\top \mathbf{x}$. Các thành phần đường chéo (diagonal component) của ma trận \mathbf{P} tạo thành ma trận $\mathbf{D} = \text{diag}(\mathbf{P})$. Lưu ý, độ phức tạp của việc nghịch đảo \mathbf{P} là $\mathcal{O}(TL^3)$, sẽ tăng nhanh khi L lớn.

Để đạt được hiệu năng cao hơn với ít số lần lặp lại, [21] đề xuất khởi tạo véc-tơ các ký hiệu đầu vào được ước lượng \mathbf{s} như trên phương trình (3.11) [6] thay vì đặt tất cả bằng 0.

$$\mathbf{s}_{in} = \mathbf{D}^{-1} \mathbf{q} = [s_0(1), s_0(2), \dots, s_0(2T)] \quad (3.11)$$

Từ véc-tơ tín hiệu thu, tín hiệu của ăng-ten/người dùng thứ j thu được bằng cách

loại bỏ tạp âm từ các ăng-ten/người dùng khác.

$$\hat{\mathbf{x}}_j = \mathbf{x} - \sum_{t=1, t \neq j}^{2T} \mathbf{h}_t \hat{s}_k(t) \quad (3.12)$$

với $\hat{\mathbf{x}}_j$ thu được, ký hiệu được gửi từ người dùng thứ j được ước lượng như sau

$$\begin{aligned} \hat{s}_{k+1}(j) &= \frac{\mathbf{h}_j^\top}{\|\mathbf{h}_j\|^2} \hat{\mathbf{x}}_j \\ &= \hat{s}_k(j) + \frac{1}{\mathbf{G}_{\mathbf{H}}(j, j)} \left(\mathbf{q}(j) - \sum_{t=1}^{2T} \mathbf{G}_{\mathbf{H}}(j, t) \hat{s}_k(t) \right) \end{aligned} \quad (3.13)$$

trong đó, \mathbf{h}_j là cột thứ j của ma trận \mathbf{H} , $\mathbf{G}_{\mathbf{H}}(i, j)$ là phần tử thứ (i, j) của ma trận $\mathbf{G}_{\mathbf{H}}$, và $\mathbf{q}(j)$ là phần tử thứ j của véc-tơ \mathbf{q} . Véc-tơ các ký hiệu ước lượng \hat{s} được cập nhật như trong thuật toán 3.1 của giải thuật ISD [21].

Thuật toán 3.1 Bộ nhận dạng Iterative Sequential [21].

Input: $\mathbf{x}, \mathbf{H}, L, T, K, \sigma^2, \mathbb{E}(\mathbf{s})$

Output: $\hat{\mathbf{s}}_{out} = \hat{\mathbf{s}}_K^{2T}$

```

1:  $\mathbf{G}_{\mathbf{H}} \leftarrow \mathbf{H}^\top \mathbf{H}$ 
2:  $\mathbf{A} \leftarrow \mathbf{G}_{\mathbf{H}} + \frac{\sigma^2}{\mathbb{E}_{\mathbf{x}}} \mathbf{I}_{2T}$ 
3:  $\mathbf{s}_0 \leftarrow \mathbf{s}_{in} = \mathbf{D}^{-1} \mathbf{q}$ 
4:
5: for  $k = 0, k < K$  do
6:   for  $j = 1, j \leq 2T$  do
7:      $\hat{s}_k(j+1) \leftarrow \hat{s}_k(j) + \frac{1}{\mathbf{G}_{\mathbf{H}}(j, j)} \left( \mathbf{q}(j) - \sum_{t=1}^{2T} \mathbf{G}_{\mathbf{H}}(j, t) \hat{s}_k(t) \right)$ 
8:
9:      $\hat{\mathbf{s}}_{k+1}^j \leftarrow [\hat{s}_{k+1}(1), \dots, \hat{s}_{k+1}(j), \hat{s}_k(j+1), \dots, \hat{s}_k(2T)]$ 
10:     $j \leftarrow j + 1$ 
11:   end for
12:    $k \leftarrow k + 1$ 
13: end for
```

Để chức minh giải thuật ISD là hiệu quả cho việc ước lượng kênh truyền, véc-tơ phần dư (lỗi) sẽ được sử dụng. Cụ thể, véc-tơ phần dư thu được sau khi khởi tạo với các giá trị \mathbf{s}_0 là

$$\mathbf{e}_0 = \mathbf{x} - \mathbf{H}\mathbf{s}_0 \quad (3.14)$$

từ đó, véc-tơ phần dư sau khi cập nhật ký hiệu thứ j tại lớp thứ k sẽ được biểu diễn như sau

$$\mathbf{e}_k^{(j)} = \mathbf{x} - \mathbf{H}\hat{\mathbf{s}}_k^j \quad (3.15)$$

thay $\mathbf{s}_k(j)$ bằng các biểu diễn hồi quy như trong giải thuật 3.1 thu được

$$\begin{aligned}
\mathbf{e}_k^{(j)} &= \mathbf{x} - \mathbf{h}_j \left(\hat{\mathbf{s}}_k(j-1) + \frac{1}{\mathbf{G}_{\mathbf{H}}(j-1, j-1)} \left(\mathbf{q}(j-1) - \sum_{t=1}^{2T} \mathbf{G}_{\mathbf{H}}(j-1, t) \hat{\mathbf{s}}_k(t) \right) \right) \\
&= \mathbf{x} - \mathbf{h}_j \left(\hat{\mathbf{s}}_k(j-1) + \frac{1}{\mathbf{h}_{j-1}^\top \mathbf{h}_{j-1}} \left(\mathbf{h}_{j-1}^\top \mathbf{x} - \sum_{t=1}^{2T} \mathbf{G}_{\mathbf{H}}(j-1, t) \hat{\mathbf{s}}_k(t) \right) \right) \\
&= \mathbf{e}_k^{(j-1)} - \mathbf{h}_j \frac{\mathbf{h}_j^\top}{\|\mathbf{h}_j\|^2} \mathbf{e}_k^{(j-1)}
\end{aligned} \tag{3.16}$$

Trong [21], Mandloi M. và các cộng sự đã chứng minh rằng $\|\mathbf{e}_k^{(j)}\|^2 < \|\mathbf{e}_k^{(j-1)}\|^2$. Điều đó chỉ ra rằng mỗi khi ký hiệu thứ j được cập nhật, véc-tơ phần dư sẽ được chiếu lên mặt phẳng ‘null’ của cột thứ j thuộc ma trận \mathbf{H} . Hay véc-tơ phần dư sẽ trực giao với \mathbf{h}_j , do đó l_2 – norm bình phương của véc-tơ lỗi sẽ giảm sau mỗi lần ký hiệu j được cập nhật cho đến khi \mathbf{e} trực giao với không gian con kéo dài bởi cột của ma trận \mathbf{H} .

3.3.2. Đề xuất mạng nơ-ron sâu ISDNN

Từ giải thuật ISD được trình bày ở trên, theo hướng tiếp cận model-driven và kỹ thuật deep unfolding, một kiến trúc mạng nơ-ron sâu có tên ISDNN (Iterative sequential deep-neural network) tương ứng được đề xuất. Đầu tiên, việc cập nhật các ký hiệu \mathbf{s} tại dòng 7 của giải thuật 3.1 được viết lại dưới dạng ma trận như sau

$$\hat{\mathbf{s}}_{k+1} = \hat{\mathbf{s}}_k + \mathbf{e}_{k+1} \tag{3.17}$$

trong đó, \mathbf{e}_{k+1} là véc-tơ phần dư cũng được viết dưới dạng ma trận là

$$\mathbf{e}_{k+1} = \mathbf{D}^{-1} \left(\mathbf{H}^\top \mathbf{x} - \mathbf{H}^\top \mathbf{H} \hat{\mathbf{s}}_k \right) \tag{3.18}$$

với \mathbf{D} được đơn giản hoá lấy ý tưởng từ bộ nhận dạng ZF khi không có thông tin về SNR tại bên thu, tức nghịch đảo của ma trận Gram $\mathbf{G}_{\mathbf{H}}$, $\mathbf{D} = \text{diag}(\mathbf{H}^\top \mathbf{H})$. Nhận thấy rằng, $\hat{\mathbf{s}}_{k+1}$ không chỉ chịu ảnh hưởng trực tiếp bởi \mathbf{e}_{k+1} mà còn tất cả các véc-tơ phần dư trước đó $\mathbf{e}_k, \mathbf{e}_{k-1}, \dots, \mathbf{e}_1$ như biểu diễn ở công thức (3.16). Do vậy, để đạt được hiệu quả cao hơn trong việc loại bỏ tạp âm từ các người dùng khác, tác giả đề xuất thêm vào các tham số học α^1 vào mỗi lớp (layer) của mạng nơ-ron.

$$\hat{\mathbf{s}}_{k+1} = \hat{\mathbf{s}}_k + \mathbf{e}_{k+1} + \alpha_k^1 \mathbf{e}_k + \alpha_{k-1}^1 \mathbf{e}_{k-1} + \dots + \alpha_1^1 \mathbf{e}_1 \tag{3.19}$$

Tuy nhiên, mối tương quan giữa các véc-tơ phần dư liên kế là lớn nhất, nên trong mạng ISDNN chỉ xem xét ảnh hưởng của \mathbf{e}_k ở lớp thứ k để đơn giản hoá mô hình. Phương

trình (3.19) trở thành

$$\mu_k = \hat{s}_k + \mathbf{e}_{k+1} + \alpha_k^1 \mathbf{e}_k \quad (3.20)$$

Thay vì gán trực tiếp $\hat{s}_{k+1} = \mu_k$, tác giả đề xuất xem xét thêm sự tương quan giữa μ_k và \hat{s}_k trước khi đưa làm đầu vào của lớp tiếp theo. Sử dụng kết hợp lồi (convex combination) [11] của \hat{s}_k và μ_k với hệ số α^2 . Do đó, \hat{s}_{k+1} chịu ảnh hưởng bởi cả \hat{s}_k và μ_k theo tỷ lệ α^2 . Trong đó, α_k^2 là tham số có thể học, $\sum_{i=k}^{k+1} \alpha_i^2 \hat{s}_i$ với $\sum_{i=k}^{k+1} \alpha_i^2 = 1$, tại mỗi lớp. Kết hợp tuyến tính của \hat{s}_k và μ_k có dạng như sau

$$\hat{s}_{k+1} = (1 - \alpha_k^2) \mu_k + \alpha_k^2 \hat{s}_k \quad (3.21)$$

Ngoài ra, để đạt được độ chính xác cao hơn ở các loại điều chế bậc cao như (16-QAM, 64-QAM, ...), véc-tơ phần dư sẽ được điều chỉnh linh hoạt hơn bằng cách thêm hai bộ biến đổi tuyến tính vào kiến trúc mạng ISDNN để cập nhật \mathbf{e}_k trước khi nhân với α_k^1 .

$$\mathbf{e}_k \leftarrow w_k^2 (w_k^1 \mathbf{e}_k + b_k^1) + b_k^2 \quad (3.22)$$

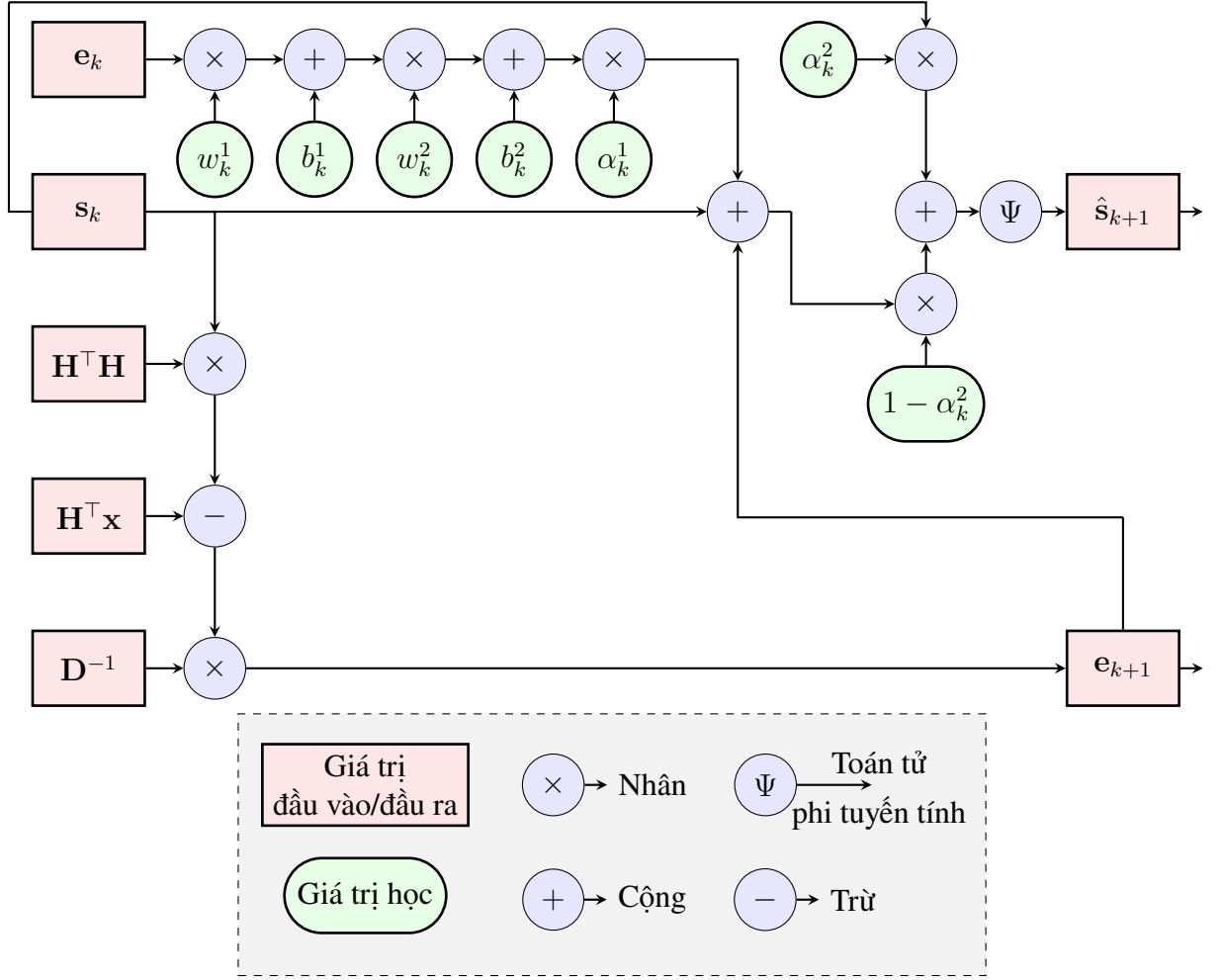
Kiến trúc cuối cùng của mạng ISDNN được đề xuất trong luận văn như trên hình 3.4. So với giải thuật ISD được đề xuất trước đó, mạng nơ-ron sâu ISDNN được đề xuất có sự cải tiến bằng việc (i) thêm véc-tơ phần dư của lớp trước đó và tham số học α^1 để ước lượng \hat{s} , (ii) tham số học α^2 được thêm vào để tăng tính chính xác của việc học, (iii) véc-tơ phần dư được đưa qua hai lớp mạng để có được tính linh hoạt cho các loại điều chế bậc cao.

Các tham số khởi tạo của mạng ISDNN được đề xuất như sau để nhanh chóng đạt được sự hội tụ [24]: $\mathbf{s}_0 = \mathbf{D}^{-1} \mathbf{q}$; α_0^1 được chọn ngẫu nhiên tiệm cận 0 ($\alpha_0^1 \approx 0$); $\alpha_0^2 = 0,5$; \mathbf{e}_0 được chọn lựa ngẫu nhiên theo phân bố chuẩn $\mathbf{e}_0 \in \mathcal{U}[0 \ 1]$. Do các đầu vào cho lớp tiếp theo \hat{s}_{k+1} cần được ánh xạ về khoảng giá trị $[-1.0 \ 1.0]$, một hàm kích hoạt (activation function) sẽ được sử dụng. Trong DL, có nhiều hàm kích hoạt được sử dụng rộng rãi như ReLu, Tanh, Sigmoid, ... như được biểu diễn trên hình 3.5. Cụ thể, trong ISDNN, tác giả lựa chọn sử dụng hàm Tanh có biểu diễn toán học như sau

$$\Psi(s) = \text{Tanh}(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}} \quad (3.23)$$

Các tham số của việc học sẽ bao gồm

$$\boldsymbol{\theta} = \{w_k^1, b_k^1, w_k^2, b_k^2, \alpha_k^1, \alpha_k^2\}_{k=1}^K \quad (3.24)$$



Hình 3.4. Kiến trúc của một lớp trong mô hình mạng nơ-ron sâu ISDNN đề xuất.

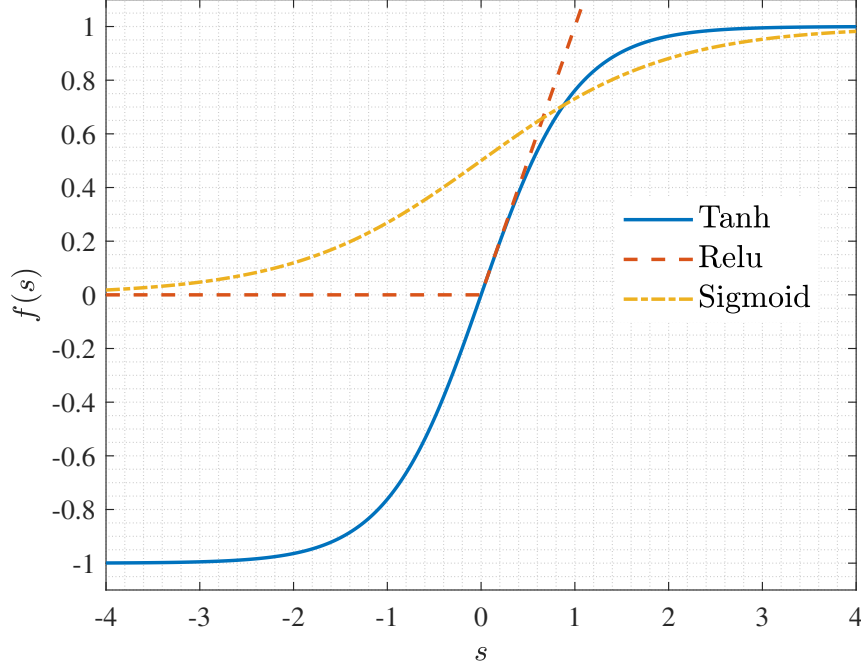
Một hàm mất mát cũng được định nghĩa như trên phương trình (3.9) để biểu diễn sự hội tụ của mô hình học ISDNN.

Bốn bước của một vòng lặp (iteration) trong quá trình học như sau

1. Khởi tạo các tham số ban đầu và véc-tơ phần dư của mạng ISDNN: $s_0, e_0, \alpha_0^1, \alpha_0^2$.
2. Bộ dữ liệu được đưa qua K lớp của mạng (forward propagation) và ước lượng sai số đầu ra $\mathcal{L}(s; \hat{s}(H, x))$.
3. Back-propagate $\mathcal{L}(s; \hat{s}(H, x))$ để thu được độ dốc (gradient).
4. Từ gradient thu được, sử dụng một thuật toán tối ưu, ví dụ là Adam [15] (xem thêm tại phụ lục A), cập nhật các $\theta = \{w_k^1, b_k^1, w_k^2, b_k^2, \alpha_k^1, \alpha_k^2\}_{k=1}^K$.

3.4. Mô phỏng và đánh giá

Trong phần này, tác giả sẽ trình bày phương pháp tạo bộ dữ liệu đào tạo cho các mô hình mạng nơ-ron sâu DetNet và ISDNN. Sau khi được đào tạo, các kết quả so sánh và đánh giá về độ chính xác, độ phức tạp, và tính chống chịu lỗi sẽ được đưa ra.



Hình 3.5. Minh họa một số hàm kích hoạt được dùng trong mô hình đề xuất.

3.4.1. Tạo bộ dữ liệu

Bảng 3.1. Các tham số mô phỏng hệ thống truyền thông không dây của mạng nơ-ron sâu ISDNN được đề xuất.

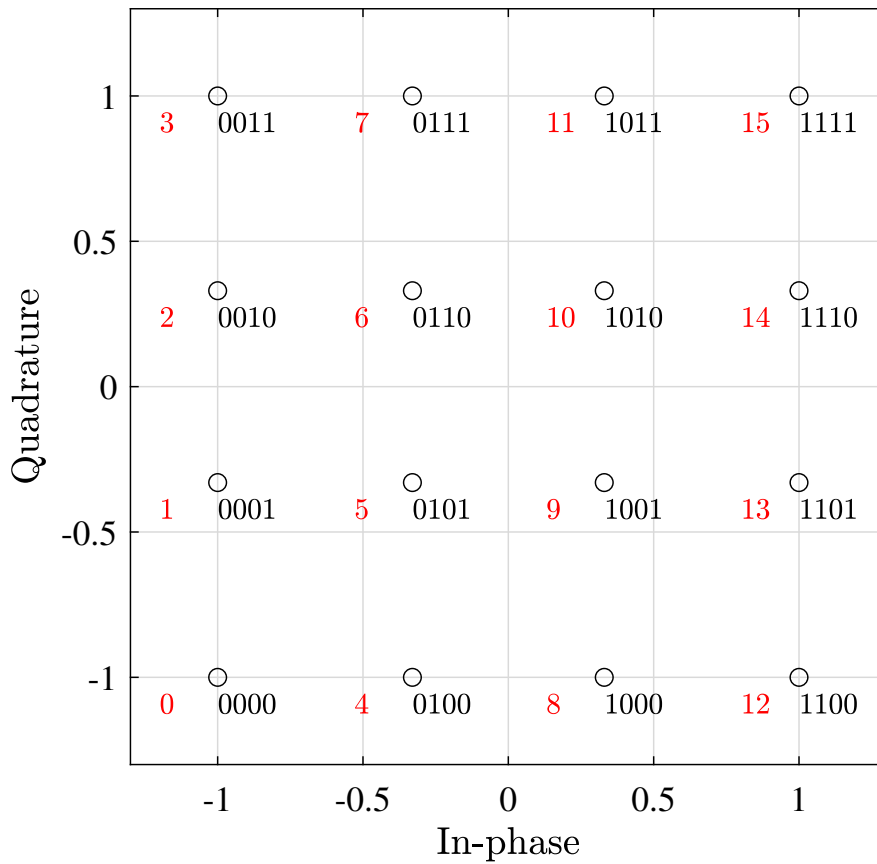
Thông số mô phỏng	Giá trị
Kích thước hệ thống mMIMO	$T = 8, L = 64$
Loại điều chế	16-QAM
Các mức SNR của dataset	$[0, 5, 10, 15, 20]$ dB
Số mẫu đào tạo	20.000
Số mẫu thử nghiệm	5.000
Số lớp mạng của DetNet	$K_{DetNet} = 4; 10$
Số lớp mạng của ISDNN	$K_{ISDNN} = 4$
Thuật toán tối ưu	Adam [15]
Giá trị khởi tạo của tốc độ học	$\delta = 0,0001$
Số vòng lặp đào tạo	20.000

Trong bảng 3.1, các tham số mô phỏng của hệ thống mMIMO cũng như kiến trúc mạng DetNet và ISDNN được đưa ra. Chi tiết, các tập dữ liệu được tạo cho việc đào tạo (training) / thử nghiệm (testing) sẽ độc lập với nhau nhưng cùng chung phân bố. Mỗi tín hiệu của bên phát s sẽ được gieo ngẫu nhiên theo phân bố đều và sử dụng chung một loại điều chế. Tuy nhiên thay vì gieo trực tiếp các ký hiệu như mô phỏng của SB-MRE là các nhóm bit $\{0, 1\}$. Tùy thuộc vào loại điều chế, mà một nhóm gồm 1, 2, 4, 8, ... bit sẽ được nhóm thành ký hiệu. Trong mô phỏng của tác giả, điều chế 16-QAM được lựa chọn, với 4 bit liên nhau sẽ được gộp lại, $u_i \in \mathbb{R}^4$, tạo thành ký hiệu s_i . Trên bảng 3.2 và

hình 3.6 là biểu diễn ánh xạ các nhóm 4 bit thành các ký hiệu s_i .

Bảng 3.2. Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM.

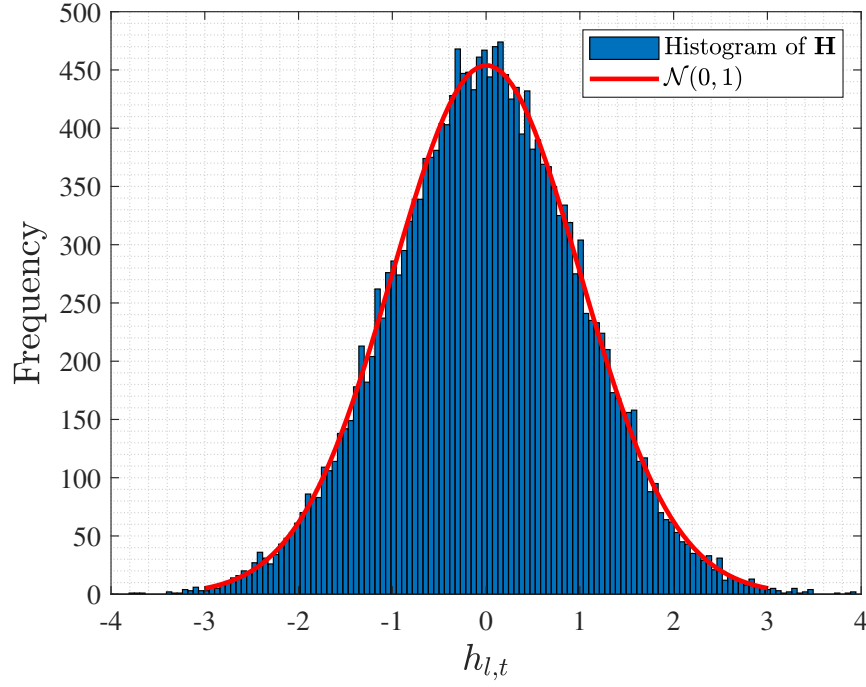
Ký hiệu	$s_0 = -1 - 1i$	$s_1 = -1 - 0,33i$	$s_2 = -1 + 0,33i$	$s_3 = -1 + 1i$
	$s_4 = -0,33 - 1i$	$s_5 = -0,33 - 0,33i$	$s_6 = -0,33 + 0,33i$	$s_7 = -0,33 + 1i$
	$s_8 = 0,33 - 1i$	$s_9 = 0,33 - 0,33i$	$s_{10} = 0,33 + 0,33i$	$s_{11} = 0,33 + 1i$
	$s_{12} = 1 - 1i$	$s_{13} = 1 - 0,33i$	$s_{14} = 1 + 0,33i$	$s_{15} = 1 + 1i$
Nhóm bit	$\mathbf{u}_0 = [0, 0, 0, 0]$	$\mathbf{u}_1 = [0, 0, 0, 1]$	$\mathbf{u}_2 = [0, 0, 1, 0]$	$\mathbf{u}_3 = [0, 0, 1, 1]$
	$\mathbf{u}_4 = [0, 1, 0, 0]$	$\mathbf{u}_5 = [0, 1, 0, 1]$	$\mathbf{u}_6 = [0, 1, 1, 0]$	$\mathbf{u}_7 = [0, 1, 1, 1]$
	$\mathbf{u}_8 = [1, 0, 0, 0]$	$\mathbf{u}_9 = [1, 0, 0, 1]$	$\mathbf{u}_{10} = [1, 0, 1, 0]$	$\mathbf{u}_{11} = [1, 0, 1, 1]$
	$\mathbf{u}_{12} = [1, 1, 0, 0]$	$\mathbf{u}_{13} = [1, 1, 0, 1]$	$\mathbf{u}_{14} = [1, 1, 1, 0]$	$\mathbf{u}_{15} = [1, 1, 1, 1]$



Hình 3.6. Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM.

Kênh truyền H lấy theo mô hình kênh Rayleigh fading, trong đó, các hệ số phức của kênh truyền được gieo ngẫu nhiên độc lập và cùng phân bố Gaussian đối xứng tròn (circularly symmetric Gaussian) với giá trị trung bình μ và độ lệch chuẩn σ^2 .

$$h_{l,t} = f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad \text{với } x \in \mathbb{R} \quad (3.25)$$



Hình 3.7. Phân bố khi gieo ngẫu nhiên của các hệ số $h_{l,t}$ trong ma trận \mathbf{H} .

Ngoài việc đào tạo mô hình dựa trên các thông tin kênh truyền \mathbf{H} chính xác, tác giả xem xét việc đào tạo ISDNN trong trường hợp thông tin \mathbf{H} không chính xác (im - imperfect) để kiểm tra khả năng chịu lỗi của mô hình đề xuất. Lý do là vì, trong các điều kiện thực tế, các ma trận đầu vào để đào tạo được đo lường không thể có được sự chính xác hoàn hảo. Hai mức sai số sẽ được xem xét đó là 1% và 5%.

$$\begin{aligned}\mathbf{H}_{im} &= \mathbf{H} \pm 0,01\mathbf{H} \\ \mathbf{H}_{im} &= \mathbf{H} \pm 0,05\mathbf{H}\end{aligned}\tag{3.26}$$

Sau khi đi qua kênh truyền \mathbf{H} , các ký hiệu sẽ được cộng thêm với AWGN ở các giá trị SNR khác nhau tính theo thang dB (decibel).

$$\text{SNR} = 10 \log \left(\frac{\mathbb{E} (\|\mathbf{H}\mathbf{s}\|_2^2)}{\mathbb{E} (\|\mathbf{w}\|_2^2)} \right) \text{ (dB)}\tag{3.27}$$

3.4.2. Đào tạo và đánh giá mô hình đề xuất

Sau khi đã tạo được các bộ dữ liệu, việc đào tạo được triển khai trên máy tính với cấu hình: vi xử lý Intel Core I9-10900, 64 GB RAM. Ngôn ngữ lập trình Python được lựa chọn để xây dựng các mô phỏng của ISDNN và DetNet. Thư viện nền tảng Pytorch được sử dụng cho ISDNN, và Tensorflow được sử dụng cho DetNet. Mã nguồn của DetNet

được sử dụng từ kho lưu trữ công khai của nhóm tác giả trên bài báo gốc tại Github¹. Sai số của các mạng nơ-ron này được đánh giá sử dụng thông số BER tương ứng là số bit ước lượng sai chia cho tổng số bit. Ở bước thử nghiệm, 100 bộ dữ liệu thử nghiệm, mỗi bộ gồm 5,000 mẫu được tạo ra, kết quả ước lượng của các mô hình sau đào tạo được tính bằng BER trung bình của 100 lần thử nghiệm.

$$\text{BER} = \frac{1}{100} \sum_{K=1}^{100} \frac{N_e}{5000} \quad (3.28)$$

Trước hết, độ phức tạp của các thuật toán sẽ được so sánh như trên bảng 3.3. Trong đó, hai bộ nhận dạng truyền thống ZF và MMSE đều có độ phức tạp $\mathcal{O}(TL^3)$ do phép nghịch đảo của ma trận \mathbf{H} với kích thước đầy đủ [27]. Tiếp đến, kiến trúc mạng DetNet cho độ phức tạp $\mathcal{O}(TL^2)$ do không phải nghịch đảo ma trận $\mathbf{G}_{\mathbf{H}}$ nên thành phần phức tạp nhất trong DetNet là các phép nhân ma trận $\mathbf{H}^T \mathbf{H}$ và $\mathbf{H}^T \mathbf{H} \mathbf{s}_k$. Cuối cùng là độ phức tạp của kiến trúc mạng ISDNN được đề xuất cũng ở mức $\mathcal{O}(TL^2)$. Dù có phép nghịch đảo ma trận \mathbf{D}^{-1} ở đầu vào, tuy nhiên như đã trình bày ở trên, ma trận \mathbf{D} chỉ gồm các phần tử trên đường chéo chính của ma trận Gram. Do vậy, việc nghịch đảo ma trận này chỉ có độ phức tạp $\mathcal{O}(TL)$, vì chỉ cần sử dụng phép biến đổi tuyến tính. Vậy nên, độ phức tạp tổng thể của ISDNN vẫn tương tự như DetNet chỉ dừng ở các phép nhân ma trận. Có thể kết luận rằng, các phương pháp sử dụng học sâu đã giảm thiểu độ phức tạp đi $\mathcal{O}(L)$ so với các bộ ước lượng tuyến tính truyền thống. Đây là khoảng cách rất lớn, vì trong các hệ mMIMO, giá trị của L có thể lên đến trên 100. So sánh riêng hai kiến trúc mạng DNN là DetNet và ISDNN, dù có chung độ phức tạp nhưng nhận thấy số lượng giá trị học của ISDNN là không đáng kể khi so sánh với DetNet. Điều này có được là do các bộ biến đổi tuyến tính (\mathbf{W} , \mathbf{b}) trong DetNet ở dưới dạng véc-tơ có kích thước lớn. Trong khi đó, ISDNN chỉ yêu cầu hai tham số học vô hướng (w , b) cho mỗi bộ biến đổi tại mỗi lớp mạng. Do vậy, chỉ 24 tham số học cần được sử dụng trong ISDNN, dẫn đến mô hình sau đào tạo chỉ có kích thước 7 KB so với 1,236 KB của DetNet với cùng số lớp mạng là 4. Đây là lợi thế rất lớn, khi kích thước nhỏ và độ phức tạp thấp giúp mô hình có thể ứng dụng trên cả các thiết bị có giá thành thấp.

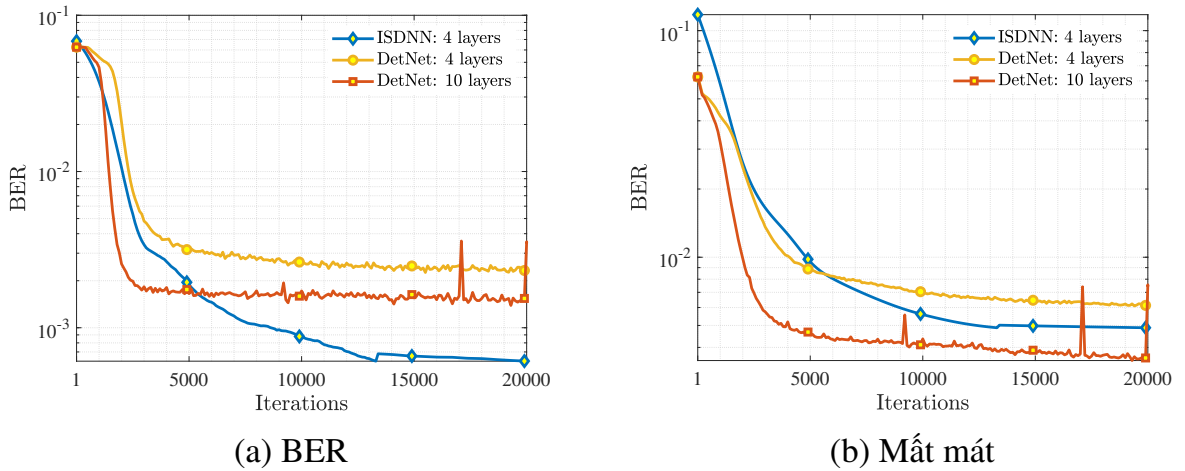
Trên hình 3.8 là quá trình đào tạo của hai mô hình mạng nơ-ron sâu ISDNN và DetNet với số lớp mạng khác nhau. Hình 3.8b xem xét về thời gian hội tụ thông qua chỉ số mất mát của véc-tơ ký hiệu gốc \mathbf{s} và véc-tơ các ký hiệu ước lượng $\hat{\mathbf{s}}$. Nhận thấy, thời gian hội tụ của mạng DetNet với 4 lớp mạng có phần nhanh hơn so với ISDNN cùng số lớp mạng. Tuy nhiên, xét về tổng thể, đầu ra hàm mất mát của ISDNN cho kết quả tốt

¹<https://github.com/neevsamuel/DeepMIMODetection>

Bảng 3.3. So sánh độ phức tạp của các thuật toán nhận dạng kênh truyền.

Bộ nhận dạng	Độ phức tạp	Số giá trị học
ZF	$\mathcal{O}(TL^3)$	
MMSE	$\mathcal{O}(TL^3)$	
DetNet: 4 layers [33]	$\mathcal{O}(TL^2)$	105.416
ISDNN	$\mathcal{O}(TL^2)$	24

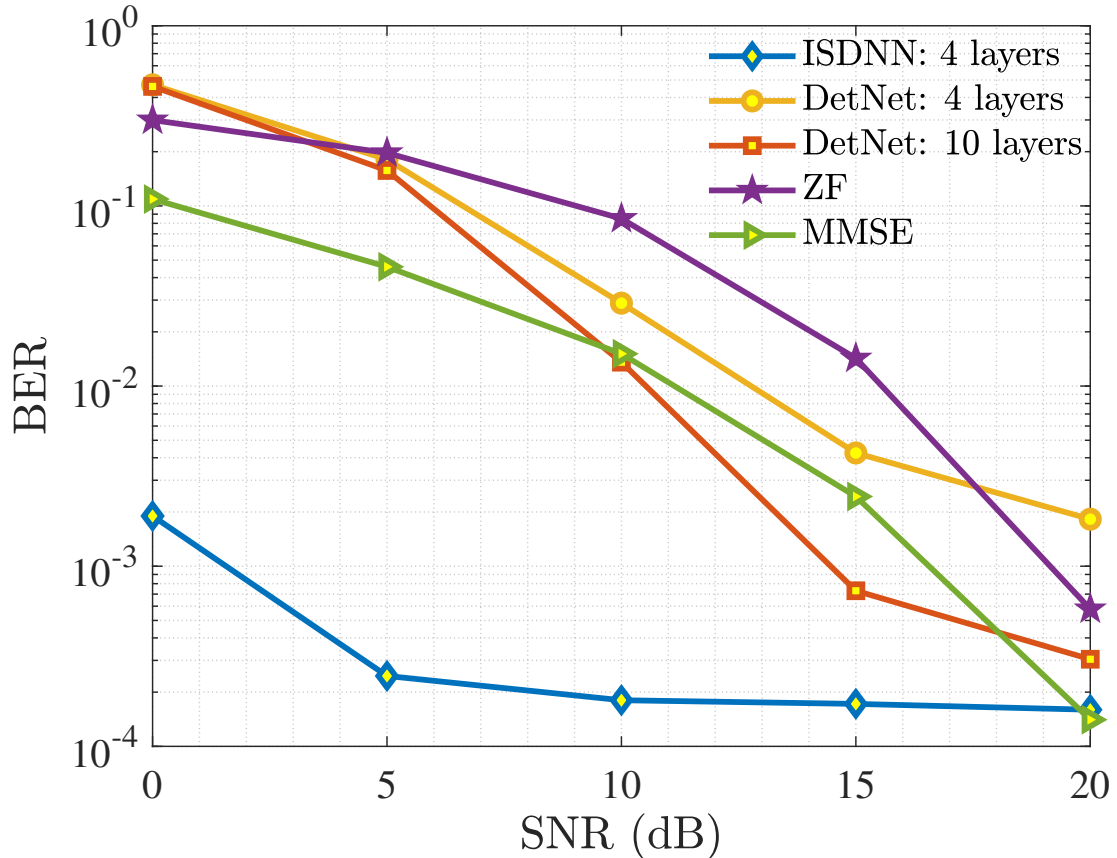
hơn so với DetNet cùng số lớp mạng dù phải cần đến vòng đào tạo thứ 12.000. Nếu tăng số lớp mạng của DetNet lên 10, do số lượng tham số học tăng lên đáng kể, thời gian hội tụ và đầu ra mất mát cuối cùng cũng cho kết quả tốt hơn ISDNN chỉ 4 lớp mạng. Tuy nhiên, đánh đổi ở đây là số lượng tham số học sẽ lên đến 316.244. Xét về sự hội tụ dựa trên độ chính xác của các mô hình như tren hình 3.8a. Trước hết, BER của ISDNN chỉ 4 lớp mạng sau 20.000 vòng đào tạo là vượt trội so với DetNet dù 4 hay 10 lớp mạng, hội tụ ở mức $BER \approx 1,6 \times 10^{-4}$. So sánh với DetNet dù với 10 lớp mạng và lượng tham số học khổng lồ cũng chỉ có đạt được sai số chưa đến 10^{-3} . Tuy nhiên, từ mô phỏng cũng cho thấy rằng, độ chính xác của DetNet cho thời gian hội tụ là nhanh hơn nhiều so với ISDNN khi chỉ cần đến khoảng 5.000 vòng đào tạo.



Hình 3.8. Sự hội tụ của quá trình đào tạo mô hình ISDNN và DetNet.

Sau quá trình đào tạo, mô hình thu được sẽ được kiểm tra trên các bộ dữ liệu thử nghiệm được tạo độc lập với tập dữ liệu huấn luyện. Kết quả thu được khi so sánh độ chính xác của các phương pháp nhận dạng kênh truyền gồm ZF, MMSE, DetNet, và ISDNN khi SNR thay đổi được biểu diễn trên hình 3.9. Trước hết, có thể kết luận, độ chính xác của mô hình đề xuất là vượt trội so với các phương pháp còn lại. Khi so sánh với hai phương pháp tuyến tính là ZF và MMSE, đường BER của ISDNN và DetNet đều cho thấy sự khác biệt, khi độ dốc của BER từ các mạng DNN giảm dần theo SNR còn ZF và MMSE thì ngược lại. Phải cần đến mức $SNR = 20$ dB, phương pháp MMSE mới

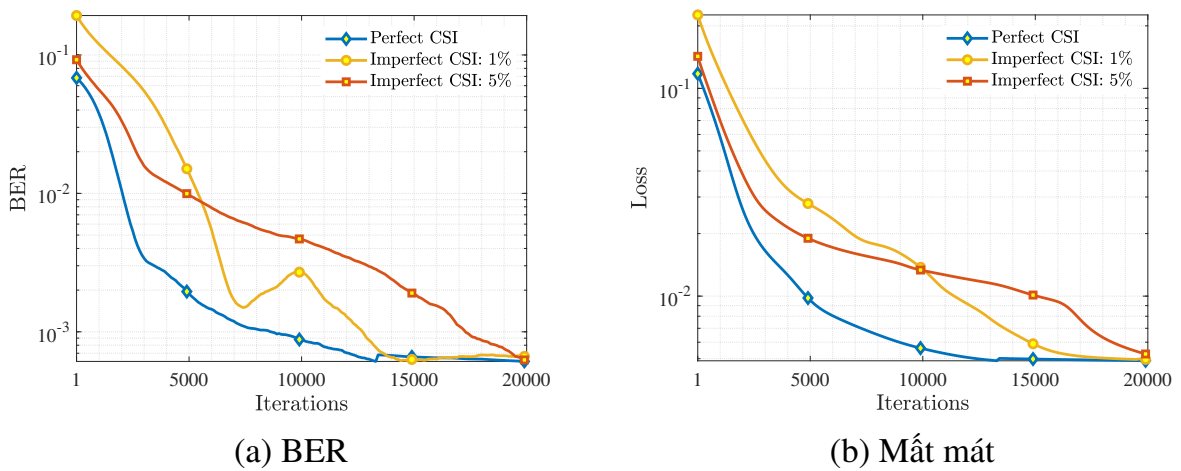
đạt đến độ chính xác của ISDNN tức $BER \approx 10^{-4}$, do giải thuật gốc ISD cũng xuất phát từ MMSE nên có thể coi đây là giá trị tối ưu của ISD. Khi so sánh với mạng nơ-ron sâu DetNet gồm 4 lớp mạng, ISDNN cũng cho độ lợi về BER đạt 10^3 tại các mức SNR thấp, và 10^1 tại SNR cao. Khi tăng số lớp của DetNet lên 10, cũng tương tự như quá trình huấn luyện, độ chính xác cũng được cải thiện, tuy nhiên dù SNR ở mức cao như 20 dB, BER của DetNet cũng chỉ tiệm cận được đến độ chính xác của ISDNN với 4 lớp mạng. Có thể rút ra nhận xét mô hình mạng ISDNN ngoài việc cho độ chính xác vượt trội so với ZF, MMSE, và DetNet còn có ưu điểm là BER không có sự biến đổi quá lớn ở các mức SNR khác nhau. Đây là đặc điểm quan trọng của mô hình đầu ra, khi tập âm / công suất phát luôn là một vấn đề mà các thể hệ mạng viễn thông thế hệ mới như 5G quan tâm. Nếu độ chính xác của việc nhận không bị ảnh hưởng nhiều bởi SNR thì mật độ bao phủ, cũng như hiệu quả về năng lượng là rõ ràng có thể nhận thấy được.



Hình 3.9. Độ chính xác của mô hình ISDNN so sánh với DetNet và các bộ nhận dạng tuyến tính.

Tiếp theo, tác giả xem xét đến tính chống chịu lỗi của mạng nơ-ron sâu ISDNN. Như đã trình bày ở trên, để có thể được áp dụng thực tế, các bộ dữ liệu cần được thu thập từ các hệ thống viễn thông thực. Tuy nhiên, sai số khi đo lường các đầu vào cho việc đào tạo là không thể tránh khỏi. Trong phần này, ma trận kênh truyền H được giả sử là có sự

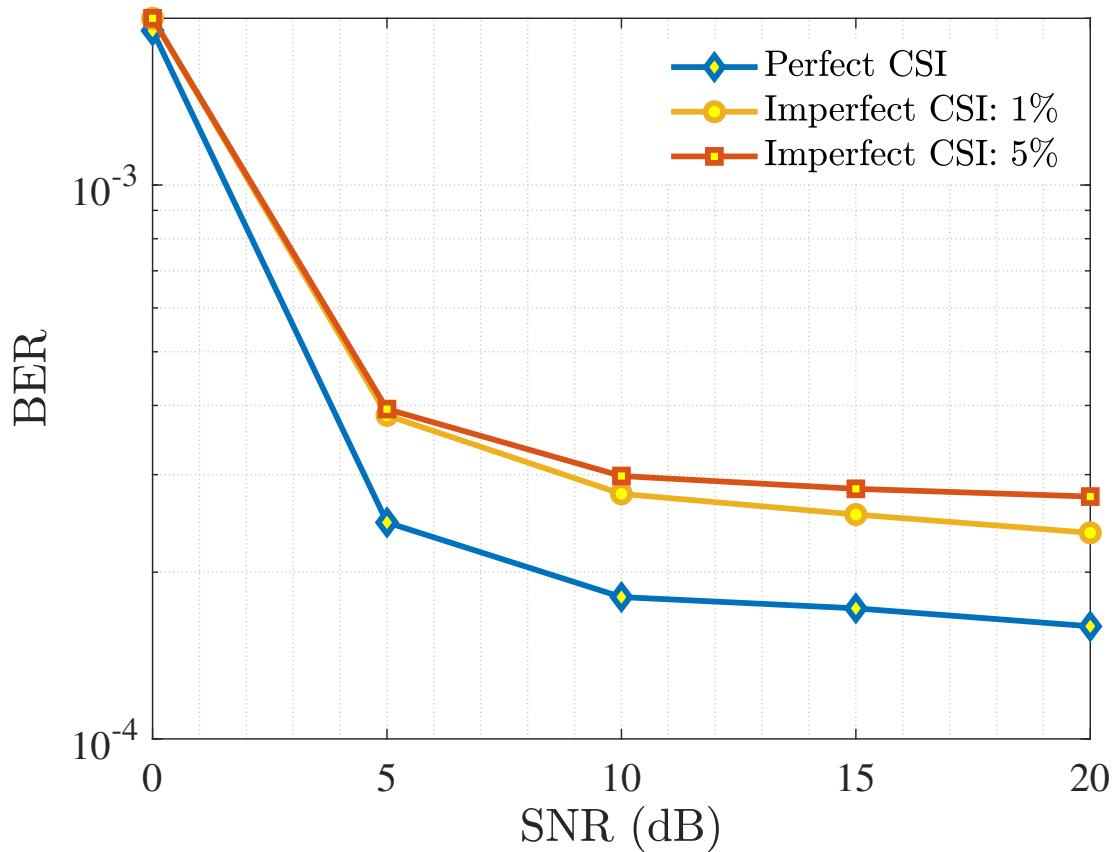
sai khác 1% và 5% so với H hoàn hảo.



Hình 3.10. Sự hội tụ của quá trình đào tạo các mô hình ISDNN với các sai số kênh truyền đầu vào khác nhau.

Trên hình 3.10 là kết quả của việc đào tạo mạng ISDNN với 3 bộ dữ liệu với các mức sai số kênh truyền khác nhau. Đầu tiên, hình 3.10b cho thấy rõ ràng sai số của dữ liệu đầu vào ảnh hưởng trực tiếp độ sự hội tụ của một mạng DNN. Trường hợp ISDNN với kênh truyền chính xác cho tốc độ hội tụ về hàm mất mát là nhanh hơn đáng kể khi so với trường hợp kênh truyền có sai số. Với sai số 1% cần đến 18.000 vòng đào tạo còn sai số 5% sau 20.000 vòng đào tạo vẫn chưa có được sự hội tụ của hàm mất mát. Tiếp theo, về độ chính xác trong quá trình đào tạo cũng cho kết quả tương tự trên hình 3.10a. Khi dữ liệu kênh truyền có sai số, đường BER trong quá trình học có sự không ổn định và cần đến từ 15.000 vòng đào tạo trở lên để hội tụ. Tuy nhiên, ở cả hàm mất mát và BER, vẫn có thể chấp nhận rằng sai số 1% của ma trận kênh truyền dù ảnh hưởng đến thời gian huấn luyện cần để hội tụ nhưng vẫn sẽ hội tụ ở giá trị tương đương với kênh truyền hoàn hảo.

Tương tự, mô hình được đào tạo ở cả ba trường hợp sẽ được đánh giá trên các bộ dữ liệu thử nghiệm, kết quả thu được như trên hình 3.11. Để nhận thấy, sự tương quan của sai số ma trận kênh truyền đầu vào với BER đầu ra của mô hình đã huấn luyện. Ở các giá trị $\text{SNR} \geq 5 \text{ dB}$, sự sai khác giữa kênh truyền chính xác và có sai số khá ổn định ở mức $\approx 1 \times 10^{-4}$. Nhưng nếu so sánh mức sai số này với các kết quả thu được trên hình 3.9, kể cả ở mức sai số 5% của ma trận kênh truyền, ISDNN vẫn cho sai số tương đương với DetNet 10 lớp mạng và vượt trội DetNet nếu chỉ 4 lớp mạng. Khi so sánh với hai bộ ước lượng tuyến tính, MMSE (ISD gốc) sẽ tốt hơn ISDNN với sai số dữ liệu 5% nếu SNR ở các giá trị $\geq 15 \text{ dB}$, ngược lại nếu SNR thấp hơn hoặc sử dụng ZF thì ISDNN vẫn sẽ có lợi thế. Từ các kết quả trên, có thể thấy sự ảnh hưởng của dữ liệu đầu vào tới ISDNN nói



Hình 3.11. Độ chính xác của mô hình ISDNN với các sai số kênh truyền đầu vào khác nhau.

riêng và các mạng DNN nói chung, tuy nhiên ở các mức sai số nhỏ, mô hình đầu ra vẫn sẽ cho độ chính xác ở mức chấp nhận được, và vẫn sẽ hơn các giải thuật tuyến tính hay mô hình DetNet. Đây chính là tiềm năng để triển khai ISDNN với các bộ dữ liệu thực, và triển khai trên các hệ thống viễn thông thực tế trong tương lai.

3.5. Kết luận chương

Trong chương này, tác giả đã trình bày khái quát về việc sử dụng DNN và mở rộng sâu cho việc nhận dạng kênh truyền. Tiếp đến, mô hình mạng nơ-ron sâu đã được đề xuất trước đây được trình bày ngắn gọn để làm cơ sở so sánh với phương pháp được tác giả đề xuất. Từ một phương pháp nhận dạng không mù với độ phức tạp thấp ISD đã được công bố trước đó, tác giả đã đề xuất mô hình mạng ISDNN mới sử dụng phương pháp mở rộng sâu. Kết quả mô phỏng đã chỉ ra hiệu năng về độ chính xác và độ phức tạp của mô hình ISDNN được đề xuất. Trước hết, về độ chính xác, ISDNN cho kết quả vượt trội so với các thuật toán nhận dạng tuyến tính không mù như ZF, MMSE, và mạng nơ-ron sâu DetNet dù với số lượng lớp mạng ít hơn. Về độ phức tạp, so với các giải thuật tuyến tính, ISDNN cho độ lợi $\mathcal{O}(L)$ tương tự như DetNet. Hơn nữa, ISDNN chỉ cần 24 tham số

học cho 4 lớp mạng khi so sánh với 105.416 tham số của mạng DetNet cùng số lớp mạng. Từ hai khía cạnh trên, kết luận mạng ISDNN đã giải quyết cả hai vấn đề là độ phức tạp, và chính xác đã được đề ra ở phần Mở đầu. Ngoài ra, để xem xét khả năng ứng dụng vào thực tế, tác giả đã xem xét hiệu suất của ISDNN nếu có sai số trong các bộ dữ liệu đầu vào, có thể xảy ra bởi sai số đo lường. Các kết quả mô phỏng chỉ ra sự ảnh hưởng của tập huấn luyện đến mô hình được đào tạo. Tuy nhiên, sai số cũng ở mức chấp nhận được và vẫn là tốt hơn nếu so sánh với các phương pháp đã kể trên.

KẾT LUẬN

Trong luận văn, tác giả tập trung giải quyết các thách thức về chi phí và độ phức tạp của các phương pháp nhận dạng trong các thể hệ mạng di động mới sử dụng mMIMO. Trước hết, một khảo sát về bốn phương pháp nhận dạng hệ thống viễn thông không dây được tác giả trình bày. Qua đó, tác giả chỉ ra sự cần thiết của việc ứng dụng chi thức mới vào bài toán nhận dạng hệ thống thông qua hai hướng tiếp cận là bán mù và sử dụng học sâu. Từ đó, trong chương 2, tác giả đã đề xuất một phương pháp nhận dạng kênh truyền cho MIMO/mMIMO đó là SB-MRE. Phương pháp sử dụng thông tin từ một số ký tự pilot và tri thức mới từ hướng tiếp cận mù là thuật toán B-MRE trước đó. Ngoài ra, tác giả cũng đề xuất giảm thiểu độ phức tạp của B-MRE nhằm giảm thiểu số lượng bộ cân bằng kênh cần ước lượng. Giải thuật được kiểm chứng qua các mô phỏng cho thấy ưu thế rõ rệt khi chỉ với một số lượng nhỏ pilot, SB-MRE đã tiệm cận và vượt qua (ở các mức SNR cao) độ chính xác của các phương pháp ZF, MMSE vốn yêu cầu đầy đủ thông tin về kênh truyền. Ngoài ra, đề xuất giảm chi phí của thành phần B-MRE cũng cho thấy kết quả ở mức chấp nhận được đặc biệt khi SNR đủ lớn để thành phần mù bắt đầu cho thấy tác dụng rõ ràng. Trong chương 3, hướng tiếp cận học sâu cũng được tác giả xem xét để nhận dạng kênh truyền cho hệ thống mMIMO. Cụ thể, một mạng ISDNN được đề xuất nhằm giảm thiểu độ phức tạp và chi phí so với thuật toán ISD gốc dựa trên MMSE. Kiến trúc mạng nơ-ron sâu được đề xuất chỉ yêu cầu 24 tham số học và 7 KB cho mô hình được đào tạo với cấu hình gồm 4 lớp mạng. Đây là số lượng rất nhỏ và hoàn toàn vượt trội khi so sánh với một mạng nơ-ron sâu khác cũng với cách tiếp cận tương tự là DetNet. Từ các kết quả mô phỏng, hiệu suất về thời gian đào tạo và độ chính xác của ISDNN cũng được kiểm chứng là vượt trội cả các phương pháp tuyến tính và mạng nơ-ron sâu DetNet. Ngoài ra, tác giả cũng xem xét đến hiệu suất của mạng nếu dữ liệu đầu vào xuất hiện sai số trong việc đo lường. Kết quả thu được cho thấy sự ảnh hưởng rõ ràng của sai số từ dữ liệu huấn luyện đến mô hình. Tuy nhiên sau khi đào tạo, độ chính xác của mô hình được đề xuất vẫn giữ được dạng gốc và vẫn có phần vượt trội so với các phương pháp khác tương tự như khi không có sai số trong dữ liệu.

Dù đạt được kết quả về mặt mô phỏng và số học, tuy nhiên, vẫn còn các điểm hạn chế có thể được cải thiện trong tương lai của luận văn. Thứ nhất là mô hình kênh truyền trong các giả thiết vẫn ở dạng đơn giản, tương tự như kênh truyền của mMIMO ở trạng thái lý tưởng khi kênh cứng. Tuy nhiên, trên thực tế, các mô hình kênh 2D, 3D có thể

cho độ lợi về độ chính xác và chi phí [34], do vậy tác giả có thể xem xét đến việc thay đổi mô hình kênh truyền trong tương lai. Thứ hai, các phương pháp như ISDNN hay DetNet được đề xuất hiện chỉ được thử nghiệm hoạt động với các kênh truyền có phân bố xác định. Tuy nhiên, điều này ít xảy ra trên thực tế, việc cải tiến, kiểm nghiệm mô hình ISDNN để đáp ứng cho trường hợp kênh truyền có phân bố thay đổi liên tục là cần thiết để đề xuất có tính ứng dụng thực tiễn hơn nữa.

**DANH MỤC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ LIÊN QUAN
ĐẾN LUẬN VĂN**

TÀI LIỆU THAM KHẢO

- [1] Abed-Meraim K., Qiu W., and Hua Y. (1997), “Blind system identification”, *Proceedings of the IEEE* 85(8), pp. 1310–1322.
- [2] Al Mtawa Y., Haque A., and Bitar B. (2019), “The mammoth internet: Are we ready?”, *IEEE Access* 7, pp. 132894–132908.
- [3] Aïssa-El-Bey A., Abed-Meraim K., and Laot C. (2011), “Adaptive blind estimation of sparse simo channels”, in *International Workshop on Systems, Signal Processing and their Applications, WOSSPA, Tipaza, Algeria*, pp. 348–351.
- [4] Bertsekas D. P. (1982), *Constrained optimization and Lagrange multiplier methods*, Academic press, New York, United States.
- [5] Chen Y., and Wainwright M. J. (2015), “Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees”, *arXiv:1509.03025* v1, pp. 1–63.
- [6] Gao X., Dai L., Ma Y., and Wang Z. (2014), “Low-complexity near-optimal signal detection for uplink large-scale mimo systems”, *Electronics Letters* 50(18), pp. 1326–1328.
- [7] Gesbert D., and Duhamel P. (1997), “Unimodal blind adaptive channel equalization: an rls implementation of the mutually referenced equalizers”, in *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, pp. 29–32.
- [8] Gesbert D., Duhamel P., and Mayrargue S. (1997), “On-line blind multichannel equalization based on mutually referenced filters”, *IEEE Transactions on Signal Processing* 45(9), pp. 2307–2317.
- [9] Gesbert D., Paulraj A., and Duhamel P. (1997), “Blind joint multiuser detection using second-order statistics and structure information”, in *Proceedings of 40th Midwest Symposium on Circuits and Systems*, Sacramento, CA, USA, pp. 1252–1255.
- [10] Giannakis G. (1997), “Filterbanks for blind channel identification and equalization”, *IEEE Signal Processing Letters* 4(6), pp. 184–187.
- [11] Hammad M. M., and Yahia M. M. (2023), *Mathematics for Machine Learning and Data Science: Optimization with Mathematica Applications*, arXiv 2302.05964, Egypt.
- [12] He H., Jin S., Wen C.-K., Gao F., Li G. Y., and Xu Z. (2019), “Model-driven deep learning for physical layer communications”, *IEEE Wireless Communications* 26(5), pp. 77–83.

- [13] Hershey J. R., Roux J. L., and Weninger F. (2014), “Deep unfolding: Model-based inspiration of novel deep architectures”, *arXiv:1409.2574* v1, pp. 1–27.
- [14] Jiang Y., Varanasi M. K., and Li J. (2011), “Performance analysis of zf and mmse equalizers for mimo systems: An in-depth study of the high snr regime”, *IEEE Transactions on Information Theory* 57(4), pp. 2008–2026.
- [15] Kingma D. P., and Ba J. (2015), “Adam: A method for stochastic optimization”, in *3rd International Conference on Learning Representations*, San Diego, CA, USA, pp. 1–15.
- [16] Ladacyia A., Belouchrani A., Abed-Meraim K., and Mokraoui A. (2019), “Semi-blind mimo-ofdm channel estimation using expectation maximisation like techniques”, *IET Communications* 13(20), pp. 3452–3462.
- [17] Ladacyia A., Mokraoui A., Abed-Meraim K., and Belouchrani A. (2017), “Performance bounds analysis for semi-blind channel estimation in mimo-ofdm communications systems”, *IEEE Transactions on Wireless Communications* 16(9), pp. 5925–5938.
- [18] Liao J., Zhao J., Gao F., and Li G. Y. (2020), “A model-driven deep learning method for massive mimo detection”, *IEEE Communications Letters* 24(8), pp. 1724–1728.
- [19] Lin X., Zhang L., and Jiang Y. (2020), “Location aided intelligent deep learning channel estimation for millimeter wave communications”, in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, Chongqing, China, pp. 489–494.
- [20] Ljung L. (1999), *System Identification: Theory for the User*, Prentice Hall information and system sciences series, Prentice Hall PTR, United States.
- [21] Mandloi M., and Bhatia V. (2017), “Low-complexity near-optimal iterative sequential detection for uplink massive mimo systems”, *IEEE Communications Letters* 21(3), pp. 568–571.
- [22] Michelusi N. (2009), *Semi-Blind Channel Estimation for LTE DownLink*, PhD thesis, Technical University Of Denmark.
- [23] Minka T. P. (2000), “Old and new matrix algebra useful for statistics”.
- [24] Narasimhan T. L., and Chockalingam A. (2014), “Channel hardening-exploiting message passing (chemp) receiver in large-scale mimo systems”, *IEEE Journal of Selected Topics in Signal Processing* 8(5), pp. 847–860.
- [25] Oh M. S., Hosseinalipour S., Kim T., Brinton C. G., and Love D. J. (2021), “Channel estimation via successive denoising in mimo ofdm systems: A reinforcement learning approach”, in *ICC 2021 - IEEE International Conference on Communications*, Montreal, QC, Canada, pp. 1–6.

- [26] Ozdemir M. K., and Arslan H. (2007), “Channel estimation for wireless ofdm systems”, *IEEE Communications Surveys & Tutorials* 9(2), pp. 18–48.
- [27] Pan V. (1992), “Complexity of computations with matrices and polynomials”, *Society for Industrial and Applied Mathematics* 34(2), pp. 225–262.
- [28] Park S., Shim B., and Choi J. W. (2015), “Iterative channel estimation using virtual pilot signals for mimo-ofdm systems”, *IEEE Transactions on Signal Processing* 63(12), pp. 3032–3045.
- [29] Rekik O., Mokraoui A., Thuy Quynh T. T., Le T.-T., and Abed-Meraim K. (2021), “Side information effect on semi-blind channel identification for mimo-ofdm communications systems”, in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp. 443–448.
- [30] Rusek F., Persson D., Lau B. K., Larsson E. G., Marzetta T. L., Edfors O., and Tufvesson F. (2013), “Scaling up mimo: Opportunities and challenges with very large arrays”, *IEEE Signal Processing Magazine* 30(1), pp. 40–60.
- [31] Samek W., Montavon G., Lapuschkin S., Anders C. J., and Müller K.-R. (2021), “Explaining deep neural networks and beyond: A review of methods and applications”, *Proceedings of the IEEE* 109(3), pp. 247–278.
- [32] Samuel N., Diskin T., and Wiesel A. (2017), “Deep mimo detection”, *arXiv:1706.01151* v1, pp. 1–5.
- [33] Samuel N., Diskin T., and Wiesel A. (2019), “Learning to detect”, *IEEE Transactions on Signal Processing* 67(10), pp. 2554–2564.
- [34] Shaik N., and Malik P. K. (2021), “A comprehensive survey 5g wireless communication systems: open issues, research challenges, channel estimation, multi carrier modulation and 5g applications”, *Multimedia Tools and Applications* 80(19), pp. 28789–28827.
- [35] Shi K., Serpedin E., and Ciblat P. (2005), “Decision-directed fine synchronization in ofdm systems”, *IEEE Transactions on Communications* 53(3), pp. 408–412.
- [36] Simeon R., Kim T., and Perrins E. (2022), “Machine learning with gaussian process regression for time-varying channel estimation”, in *ICC 2022 - IEEE International Conference on Communications*, Seoul, Korea, pp. 3400–3405.
- [37] Singh V. K., Flanagan M. F., and Cardiff B. (2019), “Generalized least squares based channel estimation for fbmc-oqam”, *IEEE Access* 7, pp. 129411–129420.
- [38] Tong L., Xu G., and Kailath T. (1994), “Blind identification and equalization based on second-order statistics: a time domain approach”, *IEEE Transactions on Information Theory* 40(2), pp. 340–349.
- [39] Treichler J., and Agee B. (1983), “A new approach to multipath correction of constant modulus signals”, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 31(2), pp. 459–472.

- [40] Tse D., and Viswanath P. (2005), *Fundamentals of Wireless Communication*, Cambridge University Press, United States.
- [41] van der Veen A.-J., and Trindade A. (2000), “Combining blind equalization with constant modulus properties”, in *34th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, pp. 1568–1572.
- [42] Vilas Boas E. C., de Figueiredo F. A., Mendes L. L., de Souza R. A. et al. (2022), “Artificial intelligence for channel estimation in multicarrier systems for b5g/6g communications: a survey”, *EURASIP Journal on Wireless Communications and Networking* 2022(1), pp. 1–63.
- [43] Wan F., Zhu W.-P., and Swamy M. N. S. (2008), “A semiblind channel estimation approach for mimo–ofdm systems”, *IEEE Transactions on Signal Processing* 56(7), pp. 2821–2834.
- [44] Wang Z., Zhao P., Qian C., and Chen S. (2016), “Location-aware channel estimation enhanced tdd based massive mimo”, *IEEE Access* 4, pp. 7828–7840.
- [45] Wisdom S., Hershey J., Le Roux J., and Watanabe S. (2016), “Deep unfolding for multichannel source separation”, in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, pp. 121–125.
- [46] Yang S., and Hanzo L. (2015), “Fifty years of mimo detection: The road to large-scale mimos”, *IEEE Communications Surveys & Tutorials* 17(4), pp. 1941–1988.
- [47] Yu C., and Xie L. (2015), “On recursive blind equalization in sensor networks”, *IEEE Transactions on Signal Processing* 63(3), pp. 662–672.
- [48] Zhang C., Patras P., and Haddadi H. (2019), “Deep learning in mobile and wireless networking: A survey”, *IEEE Communications Surveys & Tutorials* 21(3), pp. 2224–2287.

PHỤ LỤC

A. Thuật toán tối ưu Adam

Thuật toán tối ưu Adam là phương pháp tối ưu hơn so với giải thuật tối ưu giảm dần độ dốc ngẫu nhiên (SGD - Stochastic gradient descent). Adam áp dụng các tốc độ học tập thích nghi (δ - Adaptive learning rate) khác nhau cho mỗi tham số học. Điều này mang lại lợi thế lớn khi các mô hình mạng nơ-ron với kiến trúc phức tạp. Một số phần trong mạng nơ-ron nhạy cảm với sự thay đổi của trọng số theo các cách riêng biệt. Do vậy, các phần này sẽ cần tốc độ học nhỏ hơn với các vùng khác. Trong luận văn này, tác giả chỉ đưa ra một số biểu diễn toán học quan trọng của Adam, chi tiết về giải thuật tối ưu này có tại [15].

$$m_t = \beta_1 m_{t-1} - (1 - \beta_1) g_t \quad (\text{A.a})$$

$$v_t = \beta_2 v_{t-1} - (1 - \beta_2) g_t^2 \quad (\text{A.b})$$

$$\hat{m}_t = \frac{m_t}{1 + \beta_1^t} \quad (\text{A.c})$$

$$\hat{v}_t = \frac{v_t}{1 + \beta_2^t} \quad (\text{A.d})$$

$$w_t = w_{t-1} - \delta \frac{m_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (\text{A.e})$$

trong đó:

- δ : tốc độ học.
- β_1, β_2 : tỉ lệ giảm dần theo cấp số nhân cho ước lượng moment thứ nhất và hai.
- m_t : giá trị trung bình của ước lượng moment thứ nhất.
- v_t : phương sai của ước lượng moment thứ hai.
- g_t : gradient.
- \hat{m}_t : các công cụ ước lượng hiệu chỉnh bias cho moment thứ nhất.
- \hat{v}_t : các công cụ ước lượng hiệu chỉnh bias cho moment thứ hai.
- w_t : trọng số của mô hình.