

Localização Indoor com Optimização Multicritério

Pedro Miguel Rodrigues Jorge



Localização Indoor com Optimização Multicritério

Pedro Miguel Rodrigues Jorge

Dissertação para a obtenção do Grau de **Mestre em Matemática**
Área de Especialização em **Computação**

Júri

Presidente: Professor Doutor José Carlos de Gouveia Teixeira
Orientador: Professor Doutor José Luís Esteves dos Santos
Vogal: Professor Doutor João Eduardo da Silveira Gouveia

Data: Junho de 2014

Resumo

Esta tese aborda o problema de localização *Indoor* que consiste na localização de objectos ou pessoas no interior de edifícios. Utiliza a informação de dispositivos *WiFi*, onde os sistemas de localização *Outdoor*, como o *GPS*, são ineficazes devido à atenuação ou ausência de sinal nestas infraestruturas. Este é um problema actual no qual muitos recursos se têm investido com vista o seu progresso e desenvolvimento, devido à sua variedade de aplicações, principalmente em cuidados de saúde ou fins militares. Neste sentido surgiu um projecto de investigação *QREN* n° 34080 - *ADINLOC*, em co-promoção com a empresa *OnCaring*, para desenvolver uma solução de localização *Indoor* dirigida para cuidados de saúde em lares de idosos e hospitais. Esta dissertação enquadra-se no plano de trabalhos deste projecto, que possibilitou uma interacção entre a investigação e a empresa.

As estratégias de localização desenvolvidas ao longo do texto admitem que leituras de sinal recolhidas em posições próximas entre si têm valores próximos. Com base nessa ideia requerem a construção de um mapa de cobertura de sinal que contém a informação geográfica de leituras recolhidas no edifício. A utilização de métodos de optimização multicritério permite determinar as soluções do mapa de cobertura de sinal com leituras mais próximas à que foi realizada na localização que se pretende estimar. Adicionalmente, analisou-se um algoritmo da literatura baseado também nas leituras do mapa de cobertura de sinal, o algoritmo dos k vizinhos mais próximos, para comparar com as estratégias aqui desenvolvidas. Este calcula os k elementos do mapa mais próximos da leitura em termos de sinal segundo uma função utilidade.

Por fim, analisou-se o erro da localização obtida computacionalmente pelas várias estratégias relativo a 569 leituras. Esses dados indicaram que o erro médio variou entre 5.2 e 5.81 metros. Esta margem de erro permite indicar a divisão em que se encontra o objecto em cerca de 80% dos casos. Deste modo, conclui-se que este algoritmo proporciona uma boa localização grosseira, fornecendo uma boa aproximação inicial para que outros algoritmos possam realizar o seu refinamento.

Palavras Chave: Localização Indoor, Algoritmo Optimização Multicritério

Abstract

This thesis approaches the indoor location problem, which consists on estimating the position of objects or individuals inside buildings through WiFi devices, where outdoor location systems such as GPS fail due to signal attenuation or its absence inside closed infrastructures. A lot of resources have been invested nowadays for research and development of this

problem given its wide range of applications, for instance on healthcare within nursing facilities or on the army. In this sense, an investigation project QREN n. 34080 - ADINLOC in partnership with the company OnCaring emerged with the goal of developing an indoor location based solution for healthcare on seniors home and hospitals. This dissertation belongs to the work plan of that project which made possible the interaction between research and the company.

The developed strategies to locate are based on the assumption that readings collected near one another have close values as well. Bearing that in mind, they require to set up an indoor location map during an offline phase. The application of multi-criteria optimization methods finds the solutions on the signal map closer to the reading measured over the position for localisation. In addition, a known algorithm from the literature based as well on the readings from the signal map, the k nearest neighbours algorithm, was analysed in order to compare it with the algorithms developed in this thesis. It computes the k nearest elements of the signal map to the received reading according to an utility function.

At last, the error of the location given by the several strategies concerning 569 signal readings was analysed. The average error computed through that data had a range between 5.2 and 5.81 meters, a margin well adjusted to identify a division on 80% of the cases. Consequently, we can admit these strategies fit for computing a good rough location, and hence they are suitable to provide an initial approximation for other algorithms to improve.

Keywords: Indoor Location, Multi-Criteria Optimization Algorithm

Agradecimentos

Agradeço ao Departamento de Matemática da U.C. pela oportunidade de participação no projecto, ao meu orientador pelo apoio e suporte ao longo do ano, ao Engenheiro André Lemos e Jason Bolito pela ajuda nas tarefas computacionais e motivação, e à minha família pela ajuda constante.

Esta dissertação foi desenvolvida no âmbito dos trabalhos de uma bolsa de investigação para licenciados, financiada pelo Projecto ADIN-LOC: Advanced Dynamic Indoor Localization System, QREN, Agência de Inovação.



Conteúdo

1	O Problema de Localização <i>Indoor</i>	1
1.1	Introdução	1
1.2	Propagação da Força de Sinal <i>WiFi</i>	2
1.3	Mapa de Cobertura de Sinal	4
1.4	Formulações do Problema	9
2	Optimização Multicritério	11
2.1	Introdução	11
2.2	Conceitos Básicos	12
2.3	Ponderação das Soluções de Pareto	13
2.4	Aplicação ao problema de Localização <i>Indoor</i>	15
2.4.1	Localização Multi-Piso	19
3	Algoritmos	23
3.1	Introdução	23
3.2	k Vizinhos Mais Próximos	23
3.3	Optimização Multicritério	27
3.3.1	Ponderação das Soluções de Pareto	29
3.4	k Vizinhos Mais Próximos dos <i>APs</i> Visíveis na Leitura	31
3.5	k Vizinhos de Pareto Mais Próximos	32
3.6	Modificações ao Mapa de Cobertura de Sinal	32
3.7	Média Móvel	33
4	Testes Computacionais	35
4.1	Descrição dos Testes	35
4.2	Análise do Erro	36
4.2.1	Erro do Filtro de Pareto	37
4.2.2	k Vizinhos Mais Próximos	43
4.2.3	k Vizinhos de Pareto Mais Próximos	46
4.3	Resultados Multi-Piso	48
4.4	Conclusões	49
5	Trabalho Futuro	53

Capítulo 1

O Problema de Localização *Indoor*

1.1. Introdução

Cada vez mais há uma crescente preocupação para melhorar e zelar pela qualidade de vida, quer por normas governamentais sugeridas pelos Recursos Humanos ou quer pelo interesse comum de melhorar aspectos na forma como vivemos no dia a dia. Procuramos constantemente melhorias no conforto da nossa casa, na segurança dos veículos que conduzimos ou nos cuidados de saúde que nos são proporcionados.

Os sistemas de localização *Outdoor* disseminaram-se no nosso quotidiano de forma extraordinária. Por exemplo, o uso do *GPS* expandiu-se a várias áreas, sendo utilizado na aviação, na navegação marítima, nos veículos automóveis, nos passeios pedestres, na cartografia, na agricultura, etc. Esta tecnologia trouxe imensas facilidades para a nossa vida. Porém, a sua limitação surge quando se trata de localizar no interior de edifícios ou zonas com uma cobertura deficiente, consequência natural da atenuação de sinal devido a paredes, telhado ou qualquer material que interfira entre o receptor e o emissor de sinal.

Neste âmbito surge o problema de localizar dentro de infraestruturas fechadas, ou seja, o problema de localização *Indoor*. Este problema é cada vez mais importante devido à necessidade crescente de localizar no dia a dia. Por exemplo, numa instituição de cuidados de saúde, pode-se localizar uma pessoa que necessite de pedir cuidados urgentes, identificar o cuidador ou médico mais próximo, verificar se os utentes não se encontram num local indevido. Além disso, aplica-se também na construção civil para encontrar e socorrer trabalhadores em caso de acidente, ou numa empresa que constantemente necessite de procurar os operários mais próximos da área da tarefa para a sua execução.

Um sistema de localização *Indoor* é desenhado para determinar posições de determinados objectos em edifícios a partir de informações que possam ser recolhidas no seu interior. Neste contexto têm sido largamente utilizadas as estruturas de redes *WiFi*, uma vez que já se encontram presentes em muitos prédios. Assim, a locali-

zação *Indoor* irá ser realizada a partir das leituras de sinal transmitidas por *Access Points WiFi*¹ situados no seu interior.

As estratégias de localização *Indoor* primam face à localização *Outdoor*, uma vez que as características de propagação de sinal variam muito de edifício para edifício, devido a fenómenos de atenuação definidos pela sua arquitectura própria, obstáculos e tipo de material, dimensões do edifício, número de emissores, posição dos emissores, etc.

Algumas das estratégias mais utilizadas usam como descrição do sinal no edifício uma colecção de medições do sinal da rede sem fios existente a par com a sua informação espacial [CS12], denominado mapa de cobertura de sinal. O algoritmo dos k vizinhos mais próximos, que procura os k elementos do mapa de cobertura de sinal mais próximos da leitura em termos de sinal, é um exemplo clássico da utilização desse mapa [BP00]. Existem ainda outras estratégias, como por exemplo a trilateração [CS12], que recorrem a um modelo de propagação de sinal como fonte de dados para realizar a localização. Este método consiste na procura da intersecção das circunferências vizinhas de 3 *APs* nas quais é provável encontrar-se o objecto. Este método não será estudado neste trabalho, pois pertence ao plano de trabalho de outra bolsa de investigação do projecto ADINLOC em que se estuda uma técnica mais geral, a multilateração [Per14].

Num ambiente *Indoor* o sinal apresenta muita variabilidade e até na mesma posição há uma elevada dispersão das leituras registadas ao longo do tempo. Este fenómeno deve-se a ruído causado por inúmeras interferências, mudanças constantes no ambiente, atenuação dos materiais presentes, picos de sinal no emissor, até mesmo a atenuação do corpo humano quando este segura o dispositivo de leitura de sinal. Este é um factor crucial a considerar e que um algoritmo deve contemplar.

Um estudo sobre as várias abordagens de sistemas de localização *Indoor* utilizando a rede *WLAN* feito em [HLL07] indica que os mais precisos são o *EKAHAU*, com uma abordagem probabilística ou o *Robot-Based* com uma abordagem Bayesiana, apresentando erros médios na ordem dos 2 metros.

1.2. Propagação da Força de Sinal *WiFi*

A propagação de sinal da rede *WiFi* ao longo do edifício é o ponto de partida para os algoritmos de localização *Indoor*, por isso é pertinente conhecer as suas

¹Ao longo do texto um *Access Point* denotar-se-á por *AP*.

características descritas por um modelo matemático. Nesta secção descrevem-se alguns modelos de propagação de sinal.

Numa região sem obstáculos entre um AP e o receptor à distância d , com $d \geq d_0 \in \mathbb{R}^+$, o valor da potência de sinal P em dBm é dado por,

$$P = P_0 - 10\gamma \log_{10}\left(\frac{d}{d_0}\right) + \varepsilon_\sigma, \quad (1.1)$$

no qual ε_σ admite-se seguir uma distribuição normal $N(0, \sigma^2)$ que representa o erro residual [Bol13].

Porém, numa situação prática pretende-se um modelo para um ou vários pisos referente a vários APs . Neste caso é impossível não existirem obstáculos entre os APs e um dado ponto em que é registado o sinal, posto que as divisões num piso são delimitadas por paredes e a existência do tecto entre pisos. Por isso, o modelo é refeito de modo a considerar a atenuação causada pelos obstáculos existentes no edifício para cada posição. Assim, além de se considerarem todos os obstáculos existentes no edifício, em cada posição considera-se somente a atenuação causada pelos obstáculos que interferem entre essa posição e o AP . O modelo reajustado para esta situação chama-se modelo de propagação de sinal com correcção e é descrito do seguinte modo [Bol13] :

Considerem-se n tipos de obstáculos. Seja o_i o número de obstáculos do tipo i entre um AP e um receptor à distância d , com factor de atenuação λ_i , isto é, se um obstáculo do tipo i interferir entre um AP e o receptor, o sinal perde λ_i decíbel na potência recebida. A potência P , em dBm, recebida à distância $d \geq d_0$ do AP é dada por,

$$P = P_0 - 10\gamma \log_{10}\left(\frac{d}{d_0}\right) - \sum_{i=1}^n \lambda_i o_i + \varepsilon_\sigma. \quad (1.2)$$

Verifica-se que P é uma variável dependente de P_0, γ e λ_i através de uma relação linear com $\log_{10}\left(\frac{d}{d_0}\right)$ e o_i , e apresenta um erro residual normal de média 0. Conclui-se assim que este é um modelo de regressão linear em função dos parâmetros P_0, γ e λ_i . No entanto, a grande desvantagem deste modelo reside na dificuldade da contabilização automática do número de obstáculos, o_i , entre o AP e o receptor. O modelo de propagação de sinal num edifício é obtido recolhendo leituras referentes a vários APs e registando a sua posição. Procura-se, por fim, o modelo de regressão linear para cada AP descrito em (1.2) que se ajusta a esses dados.

Um dos processos para evitar o cálculo de obstáculos consiste em agrupar as recolhas de leituras em n regiões disjuntas, cada uma sem obstáculos estruturais no

seu interior. Assim, calculam-se n modelos de propagação de sinal sem correcção referentes a n regiões do edifício.

Admitindo-se que o número e tipo de obstáculos que interferem entre os APs e cada ponto de cada divisão são sempre os mesmos, verifica-se que para cada ponto da divisão se tem o mesmo factor de atenuação. Sendo o factor $\sum_{i=1}^n \lambda_i o_i$ constante para cada elemento dessa divisão, o modelo (1.2) reduz-se ao modelo de regressão linear (1.1) fazendo $P'_0 = P_0 - \sum_{i=1}^n \lambda_i o_i$.

Desta forma determinam-se n modelos de regressão linear (1.1) referentes a n divisões sem obstáculos no seu interior e a todos os APs . Se existirem m APs , o modelo da divisão i define uma função PD_i de \mathbb{R}^2 em \mathbb{R}^m que a cada ponto espacial da divisão atribui uma leitura de sinal de cada AP .

Uma característica importante a considerar num modelo de propagação de sinal (1.1), ou seja, quando não existem obstáculos, é uma leitura corresponder a uma única distância, caso se despreze o erro residual. Esta característica indica que num ambiente sem obstáculos no seu interior e com 3 APs dispostos em posições não colineares, as leituras previstas pelo modelo determinista — sem considerar o erro residual — determinam univocamente a posição numa região bidimensional.

Em suma, ainda que num ambiente com obstáculos os modelos de propagação com correcção referentes a 3 APs possam não determinar univocamente uma localização, é evidente a existência de uma forte relação entre a posição e as leituras de sinal obtidas referentes aos vários APs . Isto suporta a utilização das abordagens de localização que usam a proximidade da leitura dada em comparação com um conjunto discreto de leituras em que se conhece a posição de recolha para estimar uma localização.

1.3. Mapa de Cobertura de Sinal

Como mencionado anteriormente os algoritmos de localização *Indoor* primam em infraestruturas fechadas porque consideram uma descrição do sinal dos vários APs no seu interior. Uma destas descrições consiste no mapa de cobertura de sinal.

O mapa de cobertura de sinal contém dados de leituras recolhidas ao longo do edifício com a densidade pretendida, por exemplo, recolhas metro a metro, dois em dois metros, dez em dez centímetros, etc. Admite-se também que as leituras pertencentes ao mapa são indicadores do valor das leituras de sinal na sua vizinhança e que o mapa é representativo da propagação de sinal ao longo do piso. De facto, se uma leitura (da qual se desconhece a posição) registar para todos os APs valores

próximos de uma leitura do mapa, é provável que a sua posição (desconhecida) esteja próxima da correspondente posição (conhecida) da leitura no mapa de cobertura de sinal.

Na definição seguinte, formaliza-se a concepção de mapa de cobertura de sinal para que se possa compreender em que consiste exactamente.

Definição 1 (Mapa de Cobertura de Sinal). *Um mapa de cobertura de sinal é uma colecção de pares (p, l) , em que $p = (x, y) \in \mathbb{R}^2$ representa uma posição e $l = (l_1, \dots, l_m) \in \mathbb{R}^m$, onde l_j , $1 \leq j \leq m$, é o valor da leitura referente ao AP j registada nessa posição.*

Para facilitar, denota-se o mapa por $\Lambda = \{(p_i, l^i) \mid i = 1, \dots, M\}$, onde M representa o seu número de elementos.

Repare-se que o conceito de mapa de cobertura de sinal é extensível a \mathbb{R}^3 considerando uma terceira componente na posição. Contudo, a arquitectura da plataforma onde vão ser inseridas as estratégias de localização não se encontra preparada para isso e, por conseguinte, considerou-se apenas o caso bidimensional. Desta forma, quando se apresentam os resultados multi-piso, a terceira componente indicará apenas o piso e não propriamente a altura.

O mapa de cobertura de sinal levanta algumas questões práticas tais como, qual deverá ser o seu método de recolha? Será afectado por factores externos ou pelo decorrer do tempo? Estas questões são fundamentais porque interrogam como deve ser elaborado e os problemas a que pode estar sujeito. Por essa razão devem ser abordadas.

Em condições “ideais”, o mapa de cobertura de sinal permite dar boas estimativas da localização com base na proximidade da leitura de sinal recolhida. Contudo, em situações reais, as recolhas do mapa de cobertura estão sujeitas a ruído quando foram registadas e a disponibilidade dos APs quando se pretende realizar a localização pode ser diferente, face ao momento da recolha, pois podem não estar em funcionamento ou a cobertura variar. Além disso, como existem variações dinâmicas no ambiente ao longo do dia, em particular, no número de pessoas presentes nos vários locais, considerar-se sempre o mesmo mapa de cobertura de sinal para localizar não é a melhor opção.

Atente-se à figura 1.1 que retrata a intensidade de sinal medida na mesma posição durante um período de 24 horas numa sala de reuniões da *Critical Software*. Tendo em conta que os dados começaram a ser recolhidos às 16h15, verifica-se que no período

entre as 20h desse dia e as 8h do dia seguinte a flutuação de sinal é muito inferior. Sabe-se ainda que durante esse período não houveram reuniões pois tratava-se de um horário pós-laboral, não havendo pessoas nessa sala. Isto mostra o impacto que as pessoas têm na intensidade de sinal medida e mesmo na variabilidade de sinal.

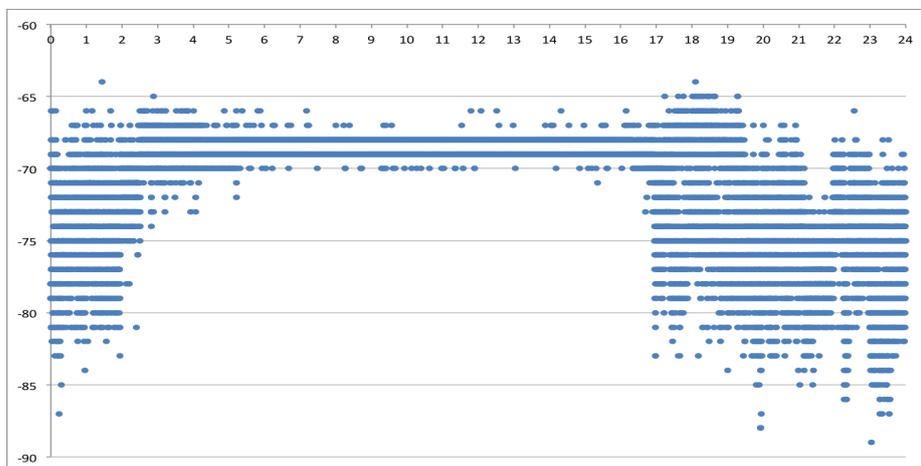


Figura 1.1: Intensidade do sinal recebido na mesma posição em função do nº de horas

Difícilmente um único mapa de cobertura de sinal seria representativo ao longo dessas 24 horas da intensidade de sinal que oscila entre os -64 dBm e os -84 dBm nessa posição.

Um método para contornar essas variações e minimizar o impacto das mudanças de sinal ao longo do tempo consiste em fazer o levantamento de vários mapas de cobertura de sinal para diferentes cenários críticos.

Para fundamentar essa ideia suponha-se que o edifício representa uma residencial. Deveria considerar-se mapas diferentes para a noite, para as horas das refeições e para a tarde, levantados num momento representativo de cada situação.

Cada mapa adequar-se-ia mais às características do sinal no seu cenário e por isso os seus dados teriam um bom impacto nas estimativas de localização. Um mapa noturno reflectiria a existência de poucas pessoas nos corredores e nos refeitórios, um mapa para a hora do pequeno almoço reflectiria as modificações em termos de sinal quando mais pessoas estariam no refeitório, na recepção e nos corredores.

Para concluir é importante mencionar que uma das etapas principais no levantamento do mapa de cobertura de sinal passa por escolher o método de recolha, por outras palavras, a disposição geográfica dos pontos a recolher e com que espaçamento.

Um método de recolha muito utilizado [CK13] consiste em percorrer caminhos em forma de S , ou seja, recolher em zigue-zague, a uma velocidade mais ou menos

constante e de seguida percorrer o mesmo caminho no sentido inverso, de forma a obter uma recolha densa de pontos e a considerar a atenuação do corpo humano conforme a sua interferência entre o receptor e os *APs*. Numa perspectiva oposta, o método adoptado nesta tese para a recolha de pontos consiste na criação de uma grelha bidimensional em que se recolhem consecutivamente várias leituras na mesma posição, o que implica a ausência de movimento durante a recolha, e pressupõe a elaboração de um novo mapa de cobertura de sinal que a cada posição tem associada a média das leituras recolhidas nessa posição com este método.

Embora a ideia crucial aplicada na construção do novo mapa através deste método de recolha tenha sido dada, esta não retrata como se calculam as médias de forma a contornar as ausências de sinal registadas numa dessas leituras, que são representadas por 0. Veja-se que tratar de forma diferente o valor 0 nesta média é obrigatório através do seguinte exemplo.

Exemplo. *Se relativamente à mesma posição fossem registadas 5 leituras com o valor $-80, -80, 0, 0, 0$ para o *AP* 1, respectivamente, e se as leituras que registassem 0 não fossem eliminadas da média, a sua média seria de $\frac{-160}{5} = -32$. Este valor representa elevada qualidade de sinal, o que não está de acordo com os dados.*

Descreve-se assim o método para o cálculo do novo mapa a partir do mapa em que foi feita a recolha das leituras.

Dada uma posição p e as leituras l^1, \dots, l^j recolhidas nessa mesma posição relativamente a m *APs*, a componente $i, 1 \leq i \leq m$, da leitura l' do novo mapa associada a essa posição é dada por,

$$l'_i = \frac{\sum_{\substack{k=1 \\ (l^k)_i \neq 0}}^j (l^k)_i}{\sum_{\substack{k=1 \\ (l^k)_i \neq 0}}^j 1}, \text{ onde } (l^k)_i \text{ é o valor da componente } i \text{ da leitura } l^k.$$

Deste modo, a leitura associada a p no novo mapa é $l' = (l'_1, \dots, l'_m)$.

Repare-se que esta média não é uniforme em todas as componentes. Sejam $1 \leq i \neq h \leq m$. O número de elementos de $\{(l^k)_i \neq 0 \mid k = 1 \dots, j\}$ pode ser diferente do número de elementos de $\{(l^k)_h \neq 0 \mid k = 1 \dots, j\}$, pois pode acontecer $(l^k)_i = 0$ e $(l^k)_h \neq 0$. Este é o motivo da média ser feita componente a componente e ignorar as componentes que não registaram sinal.

Apresentam-se duas figuras correspondentes a dois mapas de cobertura de sinal levantados na sala 4.4 do DMUC, com os dois métodos referidos, nas quais as circunferências representam onde se encontravam os *APs*.

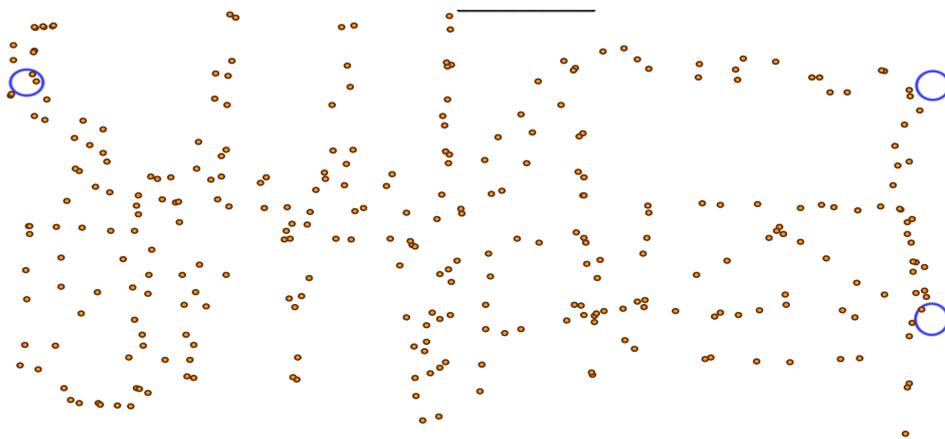


Figura 1.2: Método que percorre caminhos em forma de S

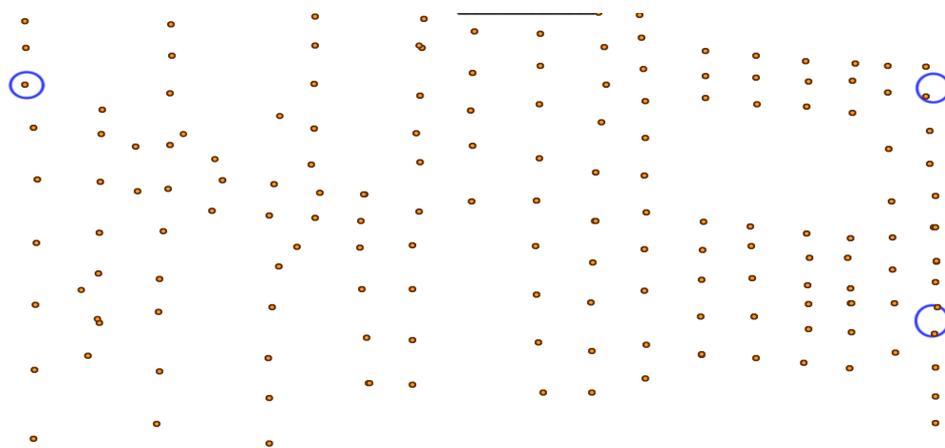


Figura 1.3: Método que recolhe uma grade de pontos

Nota-se que o mapa com estrutura de grade define uma estrutura regular, enquanto que no mapa em que se recolhe percorrendo caminhos em forma de S a disposição dos pontos é mais irregular. Além disso, a utilização das médias das leituras realizadas na mesma posição permite obter um mapa mais consistente. Por esse motivo optou-se, nesta tese, por elaborar os mapas de cobertura de sinal com o método de recolha em grade.

1.4. Formulações do Problema

Uma das abordagens empíricas ao problema de localização consiste no algoritmo dos k vizinhos mais próximos. Este algoritmo valoriza os elementos do mapa de cobertura de sinal que melhor se aproximam da leitura a localizar em termos de distância de sinal. Deste modo, precisamos de definir a função distância de sinal. Normalmente esta função é uma métrica no espaço dos sinais. Neste trabalho considerou-se como função distância de sinal a métrica euclidiana normalizada e de *Mahalonobis* descritas na página 25 da secção 3.2.

Na prática existe sempre variabilidade de sinal, assim o algoritmo dos k vizinhos mais próximos estima a localização com uma ponderação das k soluções mais próximas em vez de considerar apenas a solução mais próxima, em que é atribuído maior peso às soluções mais próximas.

Num problema de localização são fornecidos dados de leituras de sinal relativas a vários *APs*, que podem ser vistos como critérios independentes para decidir a localização ou com diferentes importâncias na sua ponderação. Neste sentido o algoritmo dos k vizinhos mais próximos consiste num problema de optimização monocritério que atribui a mesma importância a cada *AP* e ao qual tem subjacente uma função valor relativa aos vários critérios, mais especificamente os vários *APs*. Neste trabalho consideramos essencialmente a distância euclidiana normalizada e sempre que se refere este algoritmo sem especificar a métrica subjacente pressupõe-se, por defeito, que se trata desta distância.

Numa perspectiva diferente encontra-se a abordagem utilizada nesta tese que se baseia em optimização multicritério. Esta permite considerar cada *AP* como um factor diferente e definir um método de decisão de soluções para o problema com ou sem preferência relativamente a cada um desses factores, os *APs*. Esta propriedade de considerar vários factores no problema pode conseguir traduzir e captar melhor as suas características do que a versão monocritério que, em contraste, não os consegue diferenciar ou relacionar entre si por estarem agregadas a uma única função, como no algoritmo dos k vizinhos mais próximos.

A abordagem multicritério introduzida nesta tese é original, sendo um dos métodos estudados no projecto de investigação *QREN* n.º 34080 - *ADINLOC*, em copromoção com a empresa *OnCaring*, em que esta tese se enquadra.

Capítulo 2

Optimização Multicritério

2.1. Introdução

A optimização multicritério permite resolver problemas que dependem de vários critérios para decidir qual ou quais as soluções que melhor se adequam às exigências do problema. As soluções podem indicar uma certa preferência entre esses critérios ou até serem vistos como atributos com igual importância. Esta é uma área da optimização com várias aplicações, na engenharia, na medicina, na biologia e em muitas outras áreas. Na economia e finanças, em particular, surge em inúmeros problemas simples tais como pretender-se maximizar os ganhos enquanto se minimiza o risco do investimento.

Suponha-se que se pretende comprar um carro a um preço mais baixo, com menor consumo de combustível e maior potência. Este trata-se de um problema multi-objectivo em que cada critério é conflitativo, na medida que habitualmente maior potência implica maior consumo e maior preço, ou menor preço implica um motor menos potente. Pode-se então procurar carros que melhorem cada um destes aspectos e no fim comparar-se e escolher-se o que se valoriza mais no momento, se seria mais importante a potência e o consumo, ou o preço e o consumo. Enquanto se optássemos por uma abordagem de optimização monocritério não são contempladas as várias preferências possíveis e tirar partido de uma escolha *a posteriori*, porque o problema baseia-se numa única função valor que define a preferência entre soluções com base num único critério [MA04].

Na optimização multicritério decidir qual representa a melhor solução é um processo de grande importância. Este tópico deu origem a áreas que se debruçam exclusivamente sobre a vertente da tomada de decisões neste tipo de problemas.

De facto, em situações críticas, por exemplo numa aplicação financeira, o valor dado a cada critério é crucial para obter um bom resultado e podem existir diferenças na forma como se escolhem as soluções óptimas. Um caso particular seria escolher entre dois investimentos, um com menor retorno e baixo risco ou outro com

a possibilidade de elevado retorno e elevado risco de insucesso.

2.2. Conceitos B sicos

Pretende-se introduzir alguns conceitos b sicos relativos   optimizaç o multicrit rio. Um conceito essencial   o de soluç o  ptima de Pareto ou soluç o eficiente [MA04]. Este conceito permite-nos definir o que se considera como melhor soluç o num contexto em que se consideram v rios crit rios.

Definiç o 2 (Soluç o  ptima de Pareto). *Sejam $F : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, $a = (a_1, \dots, a_m)$ e $b = (b_1, \dots, b_m)$ em \mathbb{R}^m .*

- Se $\forall i \in \{1, \dots, m\} a_i \leq b_i$ e $a \neq b$ diz-se que a domina b .
- $x \in X$   soluç o  ptima de Pareto se n o existe $F(v) \in F(X)$ que domine $F(x)$.

Introduz-se um resultado interessante que apresenta uma condiç o suficiente para a exist ncia de soluç es  ptimas de Pareto nos problemas multicrit rio que ser o abordados nesta tese.

Teorema 1 (Exist ncia de Soluç es  ptimas de Pareto). *Seja $F : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Se $X \neq \emptyset$ for finito ent o existem soluç es  ptimas de Pareto.*

Demonstraç o. Considere-se a funç o $x \in X \mapsto \sum_{k=1}^m F_k(x)$. O conjunto $f(X) \subset \mathbb{R}$   finito e, por isso, tem m nimo. Seja $f(\bar{x})$ o m nimo de $f(X)$. Mostre-se que \bar{x}   soluç o  ptima de Pareto por reduç o ao absurdo.

Suponha-se que existe $F(y) \in F(X)$ que domine $F(\bar{x})$. Assim, $\forall j \in \{1, \dots, m\} F_j(y) \leq F_j(\bar{x})$, e $\exists i \in \{1, \dots, m\} F_i(y) < F_i(\bar{x})$. Obtendo-se ent o a desigualdade $\sum_{k=1}^m F_k(y) < \sum_{k=1}^m F_k(\bar{x})$. Isto implica ter-se $f(y) < f(\bar{x})$, ou seja, $f(\bar{x})$ n o   o valor m nimo de $f(X)$, o que   absurdo. \square

A seguir explica-se como se formula um problema que considera v rios factores, por outras palavras, define-se um problema multicrit rio.

Definiç o 3 (Problema de Minimizaç o). *Dado $F : X \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ formula-se o problema de optimizaç o denotado por:*

$$\min_{x \in X} F(x)$$

F diz-se a funç o objetivo e F_i , $1 \leq i \leq m$, um crit rio do problema.

Dado $x, x' \in X$, se $F_i(x) < F_i(x')$, diz-se que x melhora F no critério F_i relativamente a x' . Posto de outra forma, diz-se que x' piora F no critério F_i relativamente a x .

O conjunto $X \subseteq \mathbb{R}^n$ é a restrição imposta ao domínio do problema e denomina-se o conjunto admissível do problema.

A ordem usual \leq em \mathbb{R}^m não é total e, por isso, o conceito de mínimo tem de ser readaptado. Seja $A \subset \mathbb{R}^m$. Consideraremos o mínimo de A relativamente a uma relação de ordem parcial $R \subset A^2$, o elemento de $a \in A$ tal que $\nexists b \in A : b \neq a$ e bRa . Deste modo, atendendo à definição de solução de Pareto, determinar os mínimos de $F(X)$ com a ordem usual \leq em \mathbb{R}^m consiste em calcular as soluções óptimas de Pareto.

Como dito anteriormente a ordem em \mathbb{R}^m não é total, o que permite existirem diferentes soluções óptimas, $\{x_1, \dots, x_k\}$, do problema a verificarem $F(x_i) \neq F(x_j)$ para algum $1 \leq i \neq j \leq k$.

Para exemplificar, considere-se $x \in [0, 1] \xrightarrow{F} (x, -x)$. Verifica-se que todos os elementos em $[0, 1]$ são soluções óptimas de Pareto. Isto evidencia a possibilidade da fronteira de Pareto ser um conjunto infinito.

2.3. Ponderação das Soluções de Pareto

Na secção anterior verificou-se que os problemas terão, em geral, várias soluções óptimas. Num problema em que pretendemos indicar uma única solução, por exemplo o problema da localização *Indoor*, ter-se-á de definir um método para ponderar as diversas soluções ou, em alternativa, de preferir uma solução em particular.

Quando o conjunto das soluções de Pareto é finito, ambas as preferências referidas previamente são expressas através de uma média ponderada das várias soluções de Pareto, permitindo assim definir uma única solução.

Definição 4 (Média Ponderada). *Sejam $\omega_1, \dots, \omega_q \in \mathbb{R}_0^+$ com $\sum_{i=1}^q \omega_i > 0$ e $x_1, \dots, x_q \in \mathbb{R}^n$.*

A média ponderada de x_1, \dots, x_q é dada por

$$\bar{x} = \frac{\sum_{i=1}^q \omega_i x_i}{\sum_{i=1}^q \omega_i}.$$

Se considerarmos $1 = \omega_1 = \dots = \omega_q$, diz-se que \bar{x} é o centro de x_1, \dots, x_q .

No caso de se preferir uma soluç o  ptima x_i do conjunto de soluç es de Pareto, $\{x_j \mid j = 1, \dots, n\}$, basta considerar $\omega_i = 1$ e $\omega_j = 0$ se $j \neq i$.

Apresenta-se um teorema que apresenta uma condiç o suficiente para a optima-
lidade da m dia ponderada das soluç es de Pareto.

Teorema 2. *Se a fronteira de Pareto for um conjunto convexo a m dia ponderada de quaisquer soluç es  ptimas   tamb m soluç o  ptima de Pareto.*

Demonstraç o. Sejam S o conjunto das soluç es de Pareto convexo e $x_1, \dots, x_q \in S$.

Tem-se $\bar{x} = \sum_{i=1}^q \frac{\omega_i}{\sum_{j=1}^q \omega_j} x_i$, e por isso \bar{x}   uma combinaç o linear convexa de elementos de S . Das propriedades de um conjunto convexo, decorre que \bar{x} tamb m est  em S , ou seja,   soluç o  ptima de Pareto. \square

Para concluir, mostra-se que a m dia ponderada das soluç es de Pareto pode n o ser soluç o  ptima de Pareto. Seja F determinada por $F(x, y) = (x, y)$ e definida em $X = \{(1, 0), (0, 1), (\frac{1}{3}, \frac{1}{3}), (\frac{1}{2}, \frac{1}{3})\}$, que   um conjunto n o convexo. O problema $\min_{x \in X} F(x)$ tem como soluç es de Pareto $(1, 0), (0, 1), (\frac{1}{3}, \frac{1}{3})$. Atribua-se pesos 1, 1 e 0 a cada soluç o, respectivamente. Assim, a sua m dia ponderada   $(\frac{1}{2}, \frac{1}{2})$ e nem sequer   um elemento do conjunto admiss vel X . Note-se que, se $(\frac{1}{2}, \frac{1}{2})$ pertence-se a X , as soluç es de Pareto mantinham-se as mesmas e a m dia ponderada daria uma soluç o dominada.

Deste modo, acentua-se a import ncia de estudar o valor que se atribui a cada soluç o de modo a n o obter soluç es piores em contrapartida com escolher uma delas.

Propriedade 1. *Sejam $x^1, \dots, x^q \in \mathbb{R}^n$, $m_i = \min\{x_i^j \mid j = 1, \dots, q\}$ e $M_i = \max\{x_i^j \mid j = 1, \dots, q\}$, $i = 1, \dots, n$. A componente i , \bar{x}_i , da m dia ponderada de x^1, \dots, x^q verifica $m_i \leq \bar{x}_i \leq M_i$.*

Demonstraç o. Considere-se $1 \leq i \leq n$ e $\omega_1, \dots, \omega_q \in \mathbb{R}_0^+$ com $\sum_{j=1}^q \omega_j > 0$.

Tem-se que

$$M_i = \frac{M_i \sum_{j=1}^q \omega_j}{\sum_{j=1}^q \omega_j} = \frac{\sum_{j=1}^q \omega_j M_i}{\sum_{j=1}^q \omega_j} \geq \frac{\sum_{j=1}^q \omega_j x_i^j}{\sum_{j=1}^q \omega_j} = \bar{x}_i \geq \frac{\sum_{j=1}^q \omega_j m_i}{\sum_{j=1}^q \omega_j} = \frac{m_i \sum_{j=1}^q \omega_j}{\sum_{j=1}^q \omega_j} = m_i.$$

\square

No problema de localizaç o *Indoor*, a propriedade 1 indica que a posiç o estimada pela m dia ponderada das posiç es $\{p^j \mid j = 1, \dots, q\}$, das soluç es  ptimas de

Pareto não pertence a uma região exterior ao mapa. Desta maneira, numa *interface* gráfica se usarmos como referência uma planta do piso com os respectivos pontos do mapa de cobertura de sinal, essa localização aparecerá sempre dentro dos seus limites.

2.4. Aplicação ao problema de Localização *Indoor*

A abordagem multicritério na formulação do problema de localização parte do princípio que o mapa de cobertura de sinal é representativo da propagação de sinal da região a localizar. Deste modo, a ideia central para resolver o problema de localização consiste em procurar os elementos do mapa de sinal que melhor se aproximam da leitura que se pretende localizar, $l^* = (l_1^*, \dots, l_m^*)$, em termos de sinal, para todas as componentes, $1 \leq i \leq m$. Por outras palavras, a ideia consiste em remover os elementos para os quais existe outro que melhor se aproxime da leitura em todas as componentes, sendo os elementos não removidos as soluções de Pareto procuradas.

Para exemplificar o que se pretende com esta abordagem considere-se $l^* = (-30, -30, -30)$, a leitura realizada numa posição p^* que se pretende estimar, e as seguintes leituras do mapa de cobertura de sinal,

$$\begin{aligned} l^1 &= (p_1, \quad (-32, -32, -31)), \\ l^2 &= (p_2, \quad (-40, -40, -40)), \\ l^3 &= (p_3, \quad (-30, -31, -32)), \\ l^4 &= (p_4, \quad (-41, -30, -40)) \text{ e} \\ l^5 &= (p_5, \quad (-31, -32, -32)). \end{aligned}$$

A leitura l^2 difere mais de l^* em todos os APs do que l^1 , pelo que seria eliminada. A leitura l^5 está mais próxima de l^* no AP 1, mas está mais afastada no AP 3, relativamente a l^1 . Desta forma, nenhuma das soluções l^1 e l^5 se dominam mutuamente. Contudo a leitura l^5 é dominada por l^3 , pois l^3 aproxima-se mais de l^* em todas as componentes, enquanto que l^1 não é dominada.

Facilmente se verifica que nenhuma das leituras l^1, l^3, l^4 se aproxima mais de l^* em todas as componentes, pelo que não são dominadas mutuamente, ou seja, são soluções óptimas de Pareto.

Partindo do pressuposto que posições geográficas próximas terão leituras de sinal próximas então os elementos não dominados, em termos de proximidade componente

a componente com a leitura l^* , ser o melhores candidatos a estarem mais pr ximos da posiç o p^* .

Esta ideia   exactamente traduzida pela formulaç o (2.1) do problema multicrit rio apresentada a seguir. Dada uma leitura $l^* = (l_1^*, \dots, l_m^*)$ que se pretende localizar e um mapa de cobertura de sinal $\Lambda = \{(p_i, l^i) \mid i = 1, \dots, M\}$ da regi o em que se sup e encontrar essa leitura, a posiç o de p^* pode ser estimada a partir das soluç es de Pareto do seguinte problema:

$$\begin{aligned} \min_{i \in \{1, \dots, M\}} \quad & (l_1^i - l_1^*)^2 \\ & (l_2^i - l_2^*)^2 \\ & \vdots \\ & (l_m^i - l_m^*)^2 \end{aligned} \tag{2.1}$$

ou seja

$$\min_{i \in \{1, \dots, M\}} G_m(l^i, l^*)$$

com G_m definida por:

$$\begin{aligned} a = (a_1, \dots, a_m), b = (b_1, \dots, b_m) \in \mathbb{R}^m & \xrightarrow{G_m} ((a_1 - b_1)^2, \dots, (a_m - b_m)^2) \in \mathbb{R}^m \\ a, b \in \mathbb{R}^m & \xrightarrow{G_m} ((G_m)_1(a, b), \dots, (G_m)_m(a, b)) \in \mathbb{R}^m \end{aligned}$$

Com esta formulaç o todos os *APs* s o considerados de forma homog nea para estimar a posiç o p^* da leitura l^* , pois representam um crit rio diferente a melhorar.

Com o prop sito de obter uma  nica soluç o, como primeira abordagem considerou-se para estimativa da localizaç o o centro das soluç es de Pareto do problema (2.1). Neste caso, as soluç es contribuem equitativamente para a soluç o final, dado que se atribui o mesmo peso a cada uma.

Esta formulaç o apresenta v rias limitaç es porque, numa situaç o pr tica, a leitura a localizar recolhida pode n o registar sinal de todos os *APs* registados no mapa de cobertura de sinal. Conseqentemente, nestas condiç es a formulaç o (2.1) n o se aplica.

Nesse sentido opta-se por reajustar a leitura recolhida ao mapa de cobertura de sinal. Ou seja, considere-se uma leitura recolhida para localizar \bar{l}^* e o conjunto $J = \{j_1, \dots, j_n\} \subsetneq \{1, \dots, m\}$ dos  ndices dos *APs* que registou. Os crit rios da

leitura \bar{l}^* não coincidem com os do mapa de cobertura de sinal. Assim, opta-se por considerar uma leitura l^* no problema de optimização (2.1) definida a partir de \bar{l}^* de acordo com a situação, isto é, que indique que os *APs* em $\{1, \dots, m\} \setminus J$ não estavam visíveis. Consequentemente, para $j_s \in J$, $1 \leq s \leq h$, tem-se $l_{j_s}^* = \bar{l}_s^*$, e para $k \notin J$ tem-se $l_k^* = 0$ ¹, o que indica que o *AP* k não estava visível.

A título de exemplo considere-se um mapa de cobertura de sinal relativo aos *APs* identificados por 1, 2, 3. Se na leitura recolhida $\bar{l}^* = (l_1^*, l_3^*)$ só foram registados os *APs* 1, 3, os critérios já não coincidem com o problema (2.1), assim a leitura l^* considerada para resolver o problema é $(l_1^*, 0, l_3^*)$.

Numa análise a dados práticos verifica-se que a formulação introduzida em (2.1) devolve um grande número de soluções. De facto, basta um elemento do mapa de cobertura de sinal ter uma leitura num dos *APs* muito próxima à registada em l^* para que possivelmente seja uma solução de Pareto. Contudo, essa situação pode ocorrer devido a um pico sinal.

Ilustra-se um caso em que esta situação é crítica através da representação das diferenças quadráticas, isto é, do valor da função objectivo G em cada solução de Pareto.

$$\begin{aligned} G_6(l^1, l^*) &= (4, & 4, & 0, & 4, & 1, & 1) \\ G_6(l^2, l^*) &= (1, & 100, & 200, & 400, & 400, & 0) \\ G_6(l^3, l^*) &= (200, & 1, & 100, & 200, & 0, & 300) \\ G_6(l^4, l^*) &= (100, & 3, & 100, & 1, & 200, & 200) \end{aligned}$$

As leituras l^2, l^3 e l^4 relativamente à primeira leitura melhoram 2 critérios, enquanto que nos restantes 4 critérios pioram muito.

O ponto principal a notar é que as diferenças quadráticas de l^1 são muito pequenas e comparativamente às restantes soluções óptimas melhoram 4 dos 6 critérios, o que sugere estar mais próxima da solução pretendida. Os poucos critérios melhorados pelas restantes soluções podem ter ocorrido por um pico de sinal num dos *APs* aquando da construção do mapa de cobertura de sinal ou mesmo na recolha da leitura l^* .

Com base no exemplo anterior surge a ideia que se uma solução l^i melhora um único critério face a outra leitura l^j , então a solução l^i deve ser descartada do conjunto de soluções pretendidas.

¹O *software* disponibilizado para a recolha de sinal atribui o valor zero nas componentes da leitura em que não foi registado sinal.

Esta situação surgiu num caso prático quando se considerou como leitura l^* um elemento l de um mapa de cobertura de sinal com uma perturbação no valor das leituras de 4 dos 12 APs . Cada perturbação variou entre 1 a 2 decibel, garantindo uma diferença quadrática de l com valores em cada critério entre 0, 1 ou 4, como os que apareciam na leitura l^1 do exemplo anterior. Esta diferença quadrática indica que a posição p do elemento l daria uma boa estimativa para a posição da leitura perturbada. No entanto, após aplicado o algoritmo que calcula as soluções de Pareto, verificou-se que haviam 43 soluções do mapa de cobertura de sinal que não eram dominadas pela leitura l . Além disso, estas soluções, relativamente a l , melhoravam no máximo dois critérios e nos restantes a diferença quadrática era díspar.

Este caso prático evidencia a possibilidade de obter soluções mais concordantes com uma situação real ao impôr algumas restrições às soluções de Pareto, mais concretamente, a de cada solução melhorar mais que um critério.

No sentido de descartar possíveis más soluções ou de tentar melhorar o filtro de Pareto optou-se por estudar várias restrições às soluções em detrimento de características relevantes, como número de critérios melhorados ou APs com melhor intensidade de sinal. Repare-se que cada uma destas restrições implica uma reformulação do problema de localização.

Primeiro, em função do que foi mencionado, optou-se por excluir elementos do mapa de sinal que melhoravam no máximo n critérios, com $1 \leq n < m$. Esta restrição pode ser captada pela relação em \mathbb{R}^m assim definida:

Definição 5 (Relação de Dominância de Grau n). *Sejam $a, b \in \mathbb{R}^m$. Na relação de dominância de grau n , RD_n , com $1 \leq n \leq m$, a domina b se e só se,*

$$\#\{i \in \mathbb{N}_m \mid a_i < b_i\} \geq n \wedge \#\{i \in \mathbb{N}_m \mid a_i > b_i\} < n.$$

Posto isto, verifica-se que procurar elementos que melhoram pelo menos n critérios corresponde a procurar as soluções de Pareto relativamente à relação de dominância de grau n em vez da ordem usual de \mathbb{R}^m .

Propriedade 2. *A relação de dominância de grau 1 coincide com a relação de dominância usual em \mathbb{R}^m .*

Demonstração. Sejam $x, y \in \mathbb{R}^m$.

Tem-se $\#\{i \in \mathbb{N}_m \mid x_i < y_i\} \geq 1$ se e só se existe $j \in \{1, \dots, m\}$ com $x_j < y_j$. Por outro lado, $\#\{i \in \mathbb{N}_m \mid x_i > y_i\} < 1$, ou seja, $\#\{i \in \mathbb{N}_m \mid x_i > y_i\} = 0$, se e só

se $\forall i \in \{1, \dots, m\} x_i \leq y_i$.

Deste modo, $\#\{i \in \mathbb{N}_m \mid x_i < y_i\} \geq 1$ e $\#\{i \in \mathbb{N}_m \mid x_i > y_i\} < 1$ se e só se $\forall i \in \{1, \dots, m\} x_i \leq y_i$ e existe $j \in \{1, \dots, m\}$ com $x_j < y_j$.

Conclui-se, assim, que x domina y na relação de dominância de grau 1 se e só se x domina y na relação de dominância usual em \mathbb{R}^m . \square

A propriedade 2 evidencia que a relação de dominância usual é um caso particular da relação de dominância de grau n quando $n = 1$.

Noutra situação prática verificou-se que alguns *APs* da leitura a localizar, l^* , registavam uma intensidade de sinal baixa. Por isso estudou-se outra restrição que consiste em considerar apenas os *APs* registados em l^* com uma intensidade de sinal superior a um determinado valor $d \in \mathbb{R}$. Este valor d representa o valor mínimo aceitável de uma leitura com boa intensidade de sinal. Deste modo, as leituras com sinal fraco não têm influência no resultado, uma vez que são as mais afectadas pela variação de sinal.

Além dessas restrições, reduziu-se o problema apenas aos k , $1 \leq k \leq m$, *APs* registados na leitura, l^* , com maior intensidade de sinal. Combinou-se ainda duas destas restrições num mesmo problema: escolher as soluções não dominadas de grau n , considerando apenas os k *APs* da leitura com melhor intensidade de sinal.

Uma outra restrição consistiu em considerar somente os *APs* visíveis na leitura e escolher as soluções não dominadas de grau n . Com vista a simplificar a notação, esta formulação denomina-se filtro de Pareto aos *APs* visíveis. Note-se que na estratégia inicial, a função objectivo tinha sempre m critérios e as componentes das leituras nos *APs* não visíveis tinham o valor 0. Nesta versão irão ser considerados apenas os *APs* onde l^* registou sinal, pelo que o número de critérios pode ser inferior a m .

Para exemplificar melhor a diferença entre as duas versões considere o seguinte exemplo. Dos 5 *APs* disponíveis suponha-se que l^* apenas registou sinal nos *APs* 1, 3 e 5. Na primeira versão, a função objectivo considera os 5 critérios correspondentes aos 5 *APs* e a leitura $l^* = (l_1^*, 0, l_3^*, 0, l_5^*)$. No caso do filtro de Pareto aos *APs* visíveis, a função objectivo considera apenas 3 critérios (os *APs* visíveis) e a leitura é $l^* = (l_1^*, l_3^*, l_5^*)$.

2.4.1. Localização Multi-Piso

Os algoritmos apresentados previamente retratam somente localização referente a um piso, o que limita as suas aplicações práticas, uma vez que na maioria dos

casos pretende-se localizar num edifício com vários pisos. Para essas situações surge a necessidade de identificação do piso em que se encontra a leitura dada.

Uma abordagem intuitiva seria identificar o piso e depois tentar aplicar um algoritmo de localização a nível desse piso. Porém, pode-se tentar desenvolver um algoritmo de modo a determinar o piso e a localização conjuntamente, isto é, com os dados para determinar o piso estimar a localização da leitura sem necessitar de aplicar um algoritmo de localização referente ao piso identificado.

Uma maneira simples de abordar este problema consiste em aplicar o filtro de Pareto a um mapa de cobertura de sinal $\bar{\Lambda}$ construído a partir dos mapas de cada piso por forma a abranger toda a informação das leituras de cada piso e a ser possível compararem-se todas as leituras entre si. Assim o piso em que se prevê ter sido recolhida a leitura será o piso com mais soluções de Pareto.

Dados os mapas de cobertura de sinal $\Lambda_1, \dots, \Lambda_s$ relativos aos pisos $1, \dots, s$, respectivamente, explica-se de forma detalhada o processo de construção do mapa $\bar{\Lambda}$.

- Se o mapa do piso i tem informação dos APs $ap_1^i, \dots, ap_{i_j}^i$, o mapa $\bar{\Lambda}$ terá informação dos APs em $S = \bigcup_{i=1}^s \{ap_1^i, \dots, ap_{i_j}^i\}$.
- Dado Λ_i , cada par $(p, l) \in \Lambda_i$ corresponde a um triplo (p, l', i) em $\bar{\Lambda}$, em que l' nas componentes relativas aos APs $ap_1^i, \dots, ap_{i_j}^i$ tem o valor das componentes de l relativas a esses APs e nas componentes dos APs de $S \setminus \{ap_1^i, \dots, ap_{i_j}^i\}$ tem o valor 0.

Assim, dada uma leitura l^* e o mapa de cobertura de sinal $\bar{\Lambda}$ pretende-se localizar o piso onde foi registada l^* e a sua posição nesse piso. Deste modo, formula-se um problema multicritério relativo a l^* e $\bar{\Lambda}$. O conjunto das soluções de Pareto assim obtido $P = \{(p_1, l_1, j_1), \dots, (p_t, l_t, j_t)\}$, em que p_h indica a posição em \mathbb{R}^2 onde foi recolhida a leitura l_h e j_h o seu piso, permite determinar o piso. O piso i a que pertencerem mais soluções de Pareto será a estimativa do piso onde foi registada a leitura l^* . Os casos de empate são resolvidos com um critério de desempate, por exemplo procurar o piso que tem a solução de Pareto mais próxima da leitura em termos de sinal.

Com a intenção de estimar uma localização após se ter determinado o piso i poder-se-á optar entre aplicar um algoritmo de localização da leitura l^* ao mapa de cobertura de sinal Λ_i ou poder-se-á optar por usar as soluções óptimas de Pareto

referente a esse piso, $P_i = \{p \mid (p, l, j) \in P \wedge j = i\}$ para estimar a localização no piso.

A última opção evita cálculos no processo de determinar a localização e parte do princípio que as soluções ótimas em P_i são suficientes para estimar a localização.

É fundamental justificar em que medida P_i é suficiente para estimar a localização no piso i . Suponha-se que leitura l^* foi recolhida no piso i . Nas várias situações práticas verificou-se que, em geral, o número de soluções ótimas em P_i é inferior ao número de soluções ótimas em P'_i obtidas da aplicação do filtro de Pareto ² relativo à leitura l^* e ao mapa Λ_i . Conclui-se que algumas soluções ótimas de P'_i não eram adequadas uma vez serem dominadas por soluções de pisos diferentes. Mais ainda, verifica-se que os elementos de P_i são os que melhor se adequam ao piso i e também relativamente aos elementos de outros pisos, dado que são soluções ótimas comparativamente aos mapas de todos os pisos.

Para completar esta discussão apresenta-se um exemplo em que P_i , em certa medida, contém soluções mais ajustadas do que P'_i .

Seja l^* a leitura a localizar recolhida no piso 1 num edifício com dois pisos. Após correr o algoritmo multi-piso obteve-se

$$P = \{(p_1, 1), (p_2, 1), (p_3, 1), (p_4, 1), (p_5, 1), (p_6, 2)\},$$

que indica correctamente o piso de l^* . Com base neste resultado, este algoritmo baseia-se nas posições em $P_1 = \{p_1, p_2, p_3, p_4, p_5\}$ para estimar a localização. Por outro lado, o filtro de Pareto aplicado ao piso 1 obteve soluções

$$P'_1 = \{p_1, p_2, p_3, p_4, p_5, p, p'\}.$$

Isto mostra que p e p' são soluções dominadas por elementos do segundo piso, que não pertencem ao piso em que a leitura foi recolhida. Assim uma estimativa da localização com base nas soluções de P_1 , além de considerar as leituras ótimas do piso 1, despreza as leituras ótimas nesse piso dominadas por leituras do piso 2. Nessa medida, P_1 parece conter soluções mais ajustadas do que P'_1 pois contém somente soluções ótimas do piso correcto e para as quais não existem elementos conflituosos no piso 2, ou seja, que os dominem e pertençam ao piso errado, ao contrário do que acontece com p e p' em P'_1 .

²Identifica-se o algoritmo que calcula as soluções de Pareto por Filtro de Pareto

Capítulo 3

Algoritmos

3.1. Introdução

As abordagens ao problema de localização apresentadas nesta tese pretendem-se que sejam integradas numa solução de localização *Indoor* computacional da empresa *OnCaring*. Nessa medida foi imperativo desenvolver algoritmos que traduziam computacionalmente as várias abordagens de localização estudadas.

A plataforma disponibilizada pela *OnCaring* onde serão integrados os algoritmos de localização está desenvolvida em JAVA, uma linguagem orientada aos objetos [Sch11]. Deste modo, foi necessário estudar esta linguagem de programação de forma a adquirir competências suficientes para tornar possível a implementação dos diversos algoritmos. Os códigos escritos utilizam bibliotecas direcionadas à localização fornecidas pela *OnCaring* que permitem a aquisição de leituras de um dispositivo ANDROID e a construção de mapas de cobertura de sinal.

Antes de se prosseguir com a descrição dos algoritmos descreve-se o *input* e *output* dos algoritmos. O *input* consiste no mapa de cobertura de sinal Λ e na leitura l^* que se pretende localizar; o *output* indica a estimativa da posição p^* onde foi feita a leitura l^* . O mapa de cobertura de sinal, como já foi referido, é constituído por uma fonte de dados recolhida *a priori*, isto é, na fase *offline*. A leitura que se pretende localizar é recolhida na fase *online* com base nos sinais dos dispositivos *WiFi*, sendo realizada em períodos de, aproximadamente, 2 a 3 segundos. Para cada uma destas leituras da fase *online* aplica-se um dos algoritmos aqui desenvolvidos para indicar uma estimativa da posição (output) onde foram realizadas as leituras.

3.2. k Vizinhos Mais Próximos

Com o intuito de comparar os algoritmos que assentam em optimização multicritério com algoritmos conhecidos da literatura, optou-se por implementar o algoritmo dos k vizinhos mais próximos. Em [BP00] apresenta-se uma descrição deste algoritmo que procura os k elementos do mapa de cobertura de sinal mais próximos da

leitura a localizar relativamente a uma métrica.

Embora já tenha sido explicado de forma intuitiva, introduz-se a descrição formal do algoritmo dos k vizinhos mais próximos [BP00].

Algoritmo 1 k vizinhos mais próximos

Entrada: $((p'_i, l'_i))_{i=1, \dots, k}$ os k elementos de Λ mais próximos de uma leitura l^* em termos de uma métrica d , com $d(l'_i, l^*) \leq d(l'_{i+1}, l^*)$, $i = 1, \dots, k - 1$.

se $l'_1 = l^*$ então devolve p'_1

fim se

devolve $\frac{\sum_{i=1}^k d(l'_i, l^*)^{-1} p'_i}{\sum_{i=1}^k d(l'_i, l^*)^{-1}}$

Definição 6. *Seja $(p, l) = (p, (l_1, \dots, l_m))$ um elemento de um mapa de cobertura de sinal, $l^* = (l_1^*, \dots, l_m^*)$ uma leitura dada e $J = \{j_1, \dots, j_h\} \subset \{1, \dots, m\}$, um subconjunto de índices de APs. Define-se*

$$d_J(l, l^*) = \sqrt{\sum_{j \in J} (l_j - l_j^*)^2}$$

Atente-se a uma propriedade dada pelo seguinte teorema que indica d_J não possuir o mesmo comportamento que uma métrica.

Teorema 3. *Seja $J \subset \{1, \dots, m\}$. $d_J : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ é uma métrica se e só se $J = \{1, \dots, m\}$.*

Demonstração. Se $J = \{1, \dots, m\}$, então da sua definição verifica-se que d_J é a distância euclidiana e, portanto, uma métrica.

Suponha-se agora que $J \neq \{1, \dots, m\}$. Queremos provar que d_J não é uma métrica. Para isso basta mostrar que existem $l, l^* \in \mathbb{R}^m$ que verificam $l \neq l^*$ e $d_J(l, l^*) = 0$.

Considerem-se assim $l, l^* \in \mathbb{R}^m$ definidos por,

$$l = 0_{\mathbb{R}^m} \text{ e } l_i^* = \begin{cases} 0, & i \in J \\ 1, & i \in J^c \end{cases}$$

Obviamente se tem $l \neq l^*$ e $d_J(l, l^*) = \sqrt{\sum_{i \in J} (l_i - l_i^*)^2} = \sqrt{\sum_{i \in J} (0 - 0)^2} = 0$. □

Optou-se por implementar este algoritmo relativamente a duas métricas, a métrica euclidiana normalizada (**KVDEN**), introduzida em [Bol13], e a métrica de

Mahalonobis (KVMM). Na primeira, dada a leitura $l^* \in \mathbb{R}^m$ a localizar e uma leitura do mapa $l \in \mathbb{R}^m$, o seu valor corresponde a

$$\frac{\sqrt{\sum_{i=1}^m (l_i^* - l_i)^2}}{\#\{i \mid l_i \neq 0, i = 1, \dots, m\}}.$$

Na métrica de *Mahalonobis*, considera-se que as variáveis aleatórias que descrevem a emissão de sinal de cada AP são independentes e tem-se em conta a correlação dos dados, na medida que esta distância considera o desvio-padrão do sinal emitido por cada AP.

Teorema 4 (Distância de *Mahalonobis*). *Considere-se que (P_1, \dots, P_m) é um vector de variáveis aleatórias independentes com matriz de covariâncias Σ com $\Sigma_{ii} = \sigma_i^2 > 0$, para $i = 1, \dots, m$.*

A função d_M definida por,

$$\begin{aligned} d_M : \quad \mathbb{R}^m \times \mathbb{R}^m &\longrightarrow \mathbb{R}_0^+ \\ (x_1, \dots, x_m), (y_1, \dots, y_m) &\longmapsto \sqrt{\sum_{i=1}^m \frac{(y_i - x_i)^2}{\sigma_i^2}}, \end{aligned}$$

é uma métrica a que se chama distância de *Mahalonobis* e é induzida pelo produto interno,

$$\begin{aligned} \langle \cdot, \cdot \rangle_M : \mathbb{R}^m \times \mathbb{R}^m &\longrightarrow \mathbb{R}_0^+ \\ x, y &\longmapsto x^t \Sigma^{-1} y. \end{aligned}$$

Demonstração. Prove-se primeiro que $\langle \cdot, \cdot \rangle_M$ é um produto interno. Como (P_1, \dots, P_m) é um vector de variáveis aleatórias independentes, Σ é uma matriz diagonal. Sejam $x, y, z \in \mathbb{R}^m$ e $a \in \mathbb{R}$.

Verifica-se que $\langle x, x \rangle_M = x^t \Sigma^{-1} x = \sum_{i=1}^m \frac{x_i^2}{\sigma_i^2} \geq 0$. Tem-se também que $\langle x, x \rangle_M = 0 \iff \frac{x_i^2}{\sigma_i^2} = 0, i = 1, \dots, m \iff x_i = 0, i = 1, \dots, m \iff x = 0$.

Para mostrar a linearidade de $\langle \cdot, \cdot \rangle_M$, tem-se que $\langle x + y, z \rangle_M = (x + y)^t \Sigma^{-1} z$ e das propriedades do produto escalar decorre que $(x + y)^t \Sigma^{-1} z = x^t \Sigma^{-1} z + y^t \Sigma^{-1} z = \langle x, z \rangle_M + \langle y, z \rangle_M$. Além disso, obtém-se que $\langle ax, z \rangle_M = (ax)^t \Sigma^{-1} z = ax^t \Sigma^{-1} z = a(x^t \Sigma^{-1} z) = a \langle x, z \rangle_M$.

Para concluir, prova-se a simetria de $\langle \cdot, \cdot \rangle_M$. As seguintes igualdades são válidas, $\langle x, z \rangle_M = x^t \Sigma^{-1} z = (\Sigma^{-1} z)^t x = (z^t (\Sigma^{-1})^t) x = z^t \Sigma^{-1} x = \langle z, x \rangle_M$, pois Σ^{-1} é uma matriz simétrica.

Como $\langle \cdot, \cdot \rangle_M$ é um produto interno, $\|\cdot\|_M$, dada por $\|x\|_M = \sqrt{\langle x, x \rangle_M}$, define uma norma. Para concluir a prova, prove-se que d_M é induzido pela norma

$\|\cdot\|_M$ induzida pelo produto interno $\langle \cdot, \cdot \rangle_M$, ou seja, que se tem $d_M(x, y) = \|x - y\|_M$.

A norma $\|\cdot\|_M$ satisfaz $\|x - y\|_M = \langle x - y, x - y \rangle_M^{\frac{1}{2}} = ((x - y)^t \Sigma^{-1} (x - y))^{\frac{1}{2}}$. Como Σ^{-1} é uma matriz diagonal com $\Sigma_{ii}^{-1} = \sigma_i^{-2}, i = 1, \dots, m$, verificam-se as igualdades $(x - y)^t \Sigma^{-1} (x - y) = \sum_{i=1}^m (x_i - y_i) \Sigma_{ii}^{-1} (x_i - y_i) = \sum_{i=1}^m \frac{(x_i - y_i)^2}{\sigma_i^2}$. O que permite deduzir finalmente que $\|x - y\|_M = ((x - y)^t \Sigma^{-1} (x - y))^{\frac{1}{2}} = \sqrt{\sum_{i=1}^m \frac{(x_i - y_i)^2}{\sigma_i^2}} = d_M(x, y)$. \square

Escolheu-se a métrica euclidiana normalizada relativamente à métrica euclidiana porque, como no mapa de cobertura de sinal a ausência de sinal é representada por 0, esta métrica considera a informação acerca do número de APs visíveis na leitura, ao passo que a métrica euclidiana não.

As diferenças entre cada métrica são evidenciadas no seguinte exemplo. Sejam d a distância euclidiana e \bar{d} a distância euclidiana normalizada. Suponha-se que existem 3 APs e que as variáveis aleatórias que descrevem a emissão de sinal de cada um têm desvio-padrão 6, 5 e 3, respectivamente. Além disso, considerem-se as leituras $l^* = (-30, -30, -30)$, $l^1 = (0, -30, -31)$, $l^2 = (-60, -30, -31)$ e $l^3 = (-31, -30, -60)$.

Tem-se que $d(l^*, l^1) = d(l^*, l^2) = \sqrt{901}$ enquanto que $\bar{d}(l^*, l^2) = \frac{\sqrt{901}}{3} < \frac{\sqrt{901}}{2} = \bar{d}(l^*, l^1)$, pois a leitura l^2 apresenta 3 APs visíveis e a leitura l^1 apresenta 2 APs visíveis. Tem-se também, $d_M(l^*, l^2) = \sqrt{\frac{30^2}{36} + \frac{1}{9}} < \sqrt{\frac{1}{36} + \frac{30^2}{9}} = d_M(l^*, l^3)$, embora l^2 se obtenha de l^3 a partir de uma permutação das componentes. Isto deveu-se ao facto do sinal do AP 1 apresentar uma dispersão superior à do AP 3.

O algoritmo dos k vizinhos mais próximos com a métrica de *Mahalanobis* foi uma abordagem introduzida em [Bol13] e implementou-se com o intuito de verificar o seu desempenho na prática. Note-se que na literatura este algoritmo aparece maioritariamente associado à distância euclidiana, embora hajam outras abordagens respeitantes à métrica a utilizar [JYLS13].

Na implementação do algoritmo dos k vizinhos mais próximos com esta métrica supôs-se que as variáveis aleatórias que descrevem as leituras de sinal emitidas por cada AP são independentes entre si, ou seja, não existe condicionamento ou influência na emissão de sinal de um dado AP por parte dos restantes. Além disso, o algoritmo estima as variâncias de cada AP a partir dos valores das respeitantes leituras registadas no mapa de cobertura de sinal do seguinte modo:

Seja $\Lambda = \{(p^j, l^j) \mid j = 1, \dots, M\}$ um mapa de cobertura de sinal referente a m APs. A estimativa $\hat{\sigma}_i$ do desvio-padrão da emissão de sinal do AP i , $1 \leq i \leq m$,

determina-se a partir dos elementos com  ndices em $L_i = \{j \mid j = 1, \dots, M \wedge l_i^j \neq 0\}$. Seja $N = \#L_i$ e $\mu_i = \frac{1}{N} \sum_{j \in L_i} l_i^j$. Obt m-se finalmente $\hat{\sigma}_i^2 = \frac{1}{N-1} \sum_{j \in L_i} (l_i^j - \mu_i)^2$.

O estudo da m trica de *Mahalanobis* revela-se interessante porque, como a emiss o de sinal   uma experi ncia aleat ria, esta adopta uma perspectiva probabil stica. Repare-se que al m de considerar a informa o geom trica de cada leitura, considera tamb m o impacto da dispers o de sinal de cada *AP*.

3.3. Optimiza o Multicrit rio

Nesta sec o pretende-se enunciar as v rias estrat gias multicrit rio implementadas bem como os algoritmos centrais tais como verificar a rela o de domin ncia ou o c lculo das solu es de Pareto dada uma rela o de domin ncia. Para efeitos de contextualizar com o trabalho desenvolvido, os algoritmos s o enumerados pela ordem cronol gica em que foram concretizados.

Como introdu o descreve-se o algoritmo 2 que calcula as solu es de Pareto mas num contexto gen rico, em que a rela o de domin ncia n o precisa de ser a rela o usual em \mathbb{R}^m . No entanto deve-se reparar que o teste de domin ncia do algoritmo 2 n o   o mais eficiente e que, al m disso, se admite que na implementa o seja poss vel determinar de forma simples $\min\{n \in I \mid n > i\}$ e que as opera es $I \leftarrow I \setminus \{j\}$ possam ser executadas eficientemente.

O primeiro algoritmo implementado, o filtro de Pareto (**FP**), calcula, dada uma leitura l^* , as solu es de Pareto para a formula o do problema (2.1), a primeira abordagem multicrit rio introduzida.

De seguida passou-se   implementa o de algoritmos multicrit rio com restri es que, como foi mencionado, foram sugeridos por situa es pr ticas observadas no estudo deste problema. Estes algoritmos s o variantes do algoritmo 2 que correspondem  s formula es do problema multicrit rio (2.1) com restri es, como no n mero m nimo de crit rios a melhorar, ou no n mero de *APs* a considerar.

Em primeiro lugar implementou-se uma vers o do algoritmo 2, **FPRN**, que calcula as solu es  ptimas do mapa de cobertura de sinal quando se exige que, pelo menos, n crit rios sejam melhorados, isto  , as solu es de Pareto quando   considerada a rela o de domin ncia de grau n .

Algoritmo 2 Filtro de Pareto Genérico

Entrada: l^* , $\Lambda = \{(p^k, l^k) \mid k = 1, \dots, M\}$

$I \leftarrow \{k \in \mathbb{N} \mid 1 \leq k \leq M\}$

$i \leftarrow 0$

repetir

$i \leftarrow \min\{n \in I \mid n > i\}$

$J \leftarrow \{k \in I \mid k > i\}$

enquanto $J \neq \emptyset$ **faça**

$j \leftarrow \min J$

se $G_m(l^*, l^i)$ domina $G_m(l^*, l^j)$ **então**

$I \leftarrow I \setminus \{j\}$

$J \leftarrow J \setminus \{j\}$

senão se $G_m(l^*, l^i)$ é dominado por $G_m(l^*, l^j)$ **então**

$I \leftarrow I \setminus \{i\}$

$J \leftarrow \emptyset$

fim se

fim enquanto

até que $\#\{n \in I \mid n > i\} = 0$

devolve $\{(p^k, l^k) \in \Lambda \mid k \in I\}$ (que contém só elementos não dominados).

Posteriormente implementou-se um algoritmo, filtro de Pareto aos APs com valor superior a d (**FPAPD**), para o problema (2.1) restrito aos APs que na leitura a localizar, l^* , tivessem uma intensidade de sinal superior a um valor $d \in \mathbb{R}$. Analogamente, no sentido em que restringe APs da leitura, desenvolveu-se um algoritmo, filtro de Pareto aos k APs com maior intensidade de sinal (**FPKAP**), que considera somente os critérios dos k APs da leitura l^* com maior intensidade de sinal para calcular as soluções de Pareto.

Com o propósito de considerar duas restrições em simultâneo, desenvolveu-se um algoritmo, filtro de Pareto com duas restrições (**FPDR**), que calcula as soluções óptimas relativamente aos k APs registados na leitura com maior intensidade de sinal quando se considera a relação de dominância de grau n . Descreve-se no algoritmo 3 o método usado pelos vários algoritmos multicritério para teste de dois elementos quanto a esta relação.

Após o estudo dos algoritmos anteriores, verificou-se a partir da colecção de

leituras recolhidas que uma leitura a localizar apresenta várias componentes em que não foi registado sinal relativamente às componentes em comum com o mapa. Por isso optou-se por desenvolver um algoritmo, filtro de Pareto aos *APs* visíveis (**FPAPV**), que considera exclusivamente os critérios dos *APs* visíveis na leitura e calcula as soluções de Pareto relativamente à relação de dominância de grau n .

Algoritmo 3 Relação de Dominância de Grau n

Entrada: $v, v' \in \mathbb{R}^m$, n tal que $1 \leq n \leq m$

smaller = *greater* = 0

para $i \in \{1, \dots, m\}$ **faça**

se $v_i < v'_i$ **então**

 ++ *smaller*

senão se $v_i > v'_i$ **então**

 ++ *greater*

fim se

fim para

se *smaller* < n e *greater* $\geq n$ **então devolve** v é dominado por v'

fim se

se *greater* < n e *smaller* $\geq n$ **então devolve** v domina v'

fim se

devolve Nem v domina v' , nem v' domina v .

3.3.1. Ponderação das Soluções de Pareto

Os algoritmos descritos assentes em otimização multicritério simplesmente calculam as soluções de Pareto. Contudo, o resultado final do programa de localização que integram é uma posição. Por conseguinte, determina-se essa posição calculando uma ponderação das soluções óptimas dadas pelos algoritmos. Por isso, foi necessária a implementação de algoritmos que ponderam as soluções de Pareto.

Uma perspectiva assenta em atribuir o mesmo impacto a cada solução de Pareto na localização estimada, isto é, calcular o seu centro. Contudo, a ideia de melhorar pelo menos n critérios sugere que dadas duas soluções, a que melhorasse mais critérios relativamente à outra, deveria contribuir mais para a estimativa, isto é, ter um peso maior. Esta observação incentivou a verificar se atribuindo pesos diferentes a cada solução de Pareto se obtinha melhorias significativas nos resultados.

Propriedade 3. *Sejam* $(p^1, l^1), \dots, (p^n, l^n)$ *as soluções de Pareto da formulação (2.1)*

restrita aos critérios dos APs com índices em J relativa à leitura l^* e ao mapa de cobertura de sinal Λ referente a m APs.

Verifica-se que $d_J(l^*, l^1) = 0 \Leftrightarrow \forall i \in \{1, \dots, n\} d_J(l^*, l^i) = 0$.

Demonstração. Seja $l \in \mathbb{R}^m$. Atendendo à função objectivo G_m da formulação (2.1)

deduz-se que $d_J(l^*, l)^2 = \sum_{j \in J} (l_j^* - l_j)^2 = \sum_{j \in J} (G_m)_j(l^*, l)$.

Como $d_J(l^*, l^1)^2 = 0$ tem-se $(G_m)_j(l^*, l^1) = 0$, para todo $j \in J$.

Suponha-se que existe $i \in \{1, \dots, n\}$, com $d_J(l^*, l^i) \neq 0$. Verifica-se então que $d_J(l^*, l^i)^2 = \sum_{j \in J} (G_m)_j(l^*, l^i) > 0$. Decorre, portanto, que $\forall j \in J (G_m)_j(l^*, l^i) \geq 0$ e existe $\bar{j} \in J$ com $(G_m)_{\bar{j}}(l^*, l^i) > 0$.

Logo $\forall j \in J (G_m)_j(l^*, l^1) \leq (G_m)_j(l^*, l^i)$ e $(G_m)_{\bar{j}}(l^*, l^1) < (G_m)_{\bar{j}}(l^*, l^i)$, isto é, (p^1, l^1) domina (p^i, l^i) , o que contraria a hipótese de (p^i, l^i) ser solução de Pareto. A condição necessária (\Leftarrow) é óbvia. \square

Contudo no caso de se considerar a relação de dominância de grau n , com $n \geq 2$, esta asserção já não é válida como mostra o seguinte contra-exemplo.

Exemplo. Considere-se $l^* = (l_1, \dots, l_m)$, $l^1 = l^*$ e $l^2 = (l_1 - 1, l_2, \dots, l_m)$ as duas soluções de Pareto quando considerada a relação de dominância de grau $n \geq 2$ e $J = \{1, \dots, m\}$.

Neste caso tem-se $d_J(l^*, l^1) = 0$ e $d_J(l^*, l^2) = 1 \neq 0$. Ou seja, não se verifica a propriedade 3.

Introduz-se, finalmente, a média ponderada utilizada nos algoritmos multicritério para devolver a posição a partir das soluções de Pareto.

Definição 7 (Média Ponderada MP). *Sejam $(p^1, l^1), (p^2, l^2), \dots, (p^j, l^j)$ as soluções de Pareto da formulação (2.1) restrita aos critérios com índices em J e considerando a relação de dominância de grau n , relativa a uma leitura l^* .*

A média ponderada MP de p^1, p^2, \dots, p^j define-se do seguinte modo,

- se $d_J(l^*, l^i) = 0$ para algum i , com $1 \leq i \leq j$, MP é o centro dos elementos de

$$\{p^i \mid 1 \leq i \leq j \wedge d_J(l^*, l^i) = 0\};$$

- caso contrário, MP é dada por $\frac{\sum_{i=1}^j \omega_i p^i}{\sum_{i=1}^j \omega_i}$, com $\omega_i = \frac{1}{d_J(l^*, l^i)}$, $i = 1, \dots, j$.

A atribuição dos pesos em MP tem por base a premissa referida na secção 1.3 que a posições geográficas próximas devem corresponder leituras próximas. Note-se que no caso de $d_J(l^*, l^i) = 0$ para algum i , a média ponderada MP atribui a todas as soluções (p, l) que verificam $d_J(l^*, l) = 0$ o mesmo peso, pois estas soluções do ponto de vista multicritério e relativamente à proximidade com a leitura l^* recolhida não apresentam preferência entre si. Quando $\forall i \ d_J(l^*, l^i) \neq 0$, a média ponderada MP atribui peso maior às soluções que melhor se aproximem da leitura l^* em termos de sinal relativamente aos APs considerados no cálculo destas soluções.

A implementação do algoritmo que calcula esta média ponderada foi desenvolvido para comparar o desempenho do filtro de Pareto quando atribui o mesmo impacto a todas as soluções óptimas para estimar a localização, isto é, quando calcula o seu centro, com o caso de valorizar soluções conforme as suas características. No caso desta média a característica valorizada é a proximidade da solução à leitura a localizar em termos de sinal relativamente aos APs considerados pelo algoritmo.

3.4. k Vizinhos Mais Próximos dos APs Visíveis na Leitura

Com o intuito de comparar o algoritmo **FPAPV** com os k vizinhos mais próximos desenvolveu-se uma versão do algoritmo dos k vizinhos mais próximos de forma análoga.

Definição 8. *Seja $l = (l_1, \dots, l_m) \in \mathbb{R}^m$ e $J = \{j_1, \dots, j_i\} \subset \{1, \dots, m\}$, um subconjunto de índices de APs , define-se C_J por,*

$$C_J(l) = (l_{j_1}, l_{j_2}, \dots, l_{j_i}) \in \mathbb{R}^i$$

Repare-se que, se l^* é a leitura a localizar e J o conjunto dos índices dos APs visíveis em l^* , $C_J(l^*)$ corresponde a um vector com as componentes de l^* em J .

Seja $\Lambda = \{(p^i, l^i) \mid i = 1, \dots, M\}$ um mapa de cobertura de sinal, l^* a leitura a localizar e J o conjunto dos índices dos APs visíveis em l^* . O algoritmo dos k vizinhos mais próximos dos APs visíveis (**KVAPV**), difere da versão original do algoritmo dos k vizinhos mais próximos, pois em vez de calcular os k vizinhos mais próximos relativos a l^* e Λ , calcula os k vizinhos mais próximos da leitura $C_J(l^*)$ considerando a distância euclidiana normalizada, relativamente ao mapa $\Lambda' = \{(p^i, C_J(l^i)) \mid i = 1, \dots, M\}$ dos quais estima a posição.

A analogia com o algoritmo **FPAPV** encontra-se no facto de se considerarem somente os APs do mapa que se encontram visíveis na leitura a localizar.

3.5. k Vizinhos de Pareto Mais Próximos

Com os algoritmos apresentados surgiu a ideia de desenvolver um algoritmo híbrido, os k vizinhos de Pareto mais próximos (**KVP**). Este algoritmo teria como objectivo verificar se estimar uma localização a partir da informação geométrica das soluções de Pareto seria o suficiente para estimar uma boa localização.

A ideia base consiste em calcular os k vizinhos mais próximos das soluções de Pareto do problema na formulação (2.1). Desta forma, através da análise do erro verificar-se-ia as melhorias, caso existissem, de aplicar o algoritmo dos k vizinhos mais próximos simplesmente à fronteira de Pareto em vez do mapa completo.

Por fim, descreve-se os k vizinhos de Pareto mais próximos de forma genérica.

Algoritmo 4 k Vizinhos de Pareto Mais Próximos

Entrada: l^* , $\Lambda = \{(p_i, l_i) \mid i = 1, \dots, M\}$

Calcular $S = \{s_1, \dots, s_n\} \subset \Lambda$, as soluções de Pareto obtidas com o algoritmo **FP** relativo a l^* e Λ ;

Calcular os k vizinhos $(p_{v_1}, l_{v_1}), \dots, (p_{v_k}, l_{v_k})$ em S mais próximos de l^* , com $d(l_{v_i}, l^*) \leq d(l_{v_{i+1}}, l^*), i = 1, \dots, k - 1$;

se $l_{v_1} = l^*$ então devolve p_{v_1}

senão devolve $\frac{\sum_{i=1}^k d(l_{v_i}, l^*)^{-1} p_{v_i}}{\sum_{i=1}^k d(l_{v_i}, l^*)^{-1}}$

fim se

3.6. Modificações ao Mapa de Cobertura de Sinal

Quando um AP não está visível numa leitura do mapa de cobertura de sinal a aplicação disponibilizada pela *OnCaring* regista o valor 0. Contudo, esse valor pode originar conflitos no problema de optimização. De facto, suponha-se que a leitura recolhida l^* registou -40 dBm num AP i e que um elemento (p, l) nesse AP registou o valor 0 por estar, por exemplo, muito afastado da posição onde se encontra esse AP . Assim (p, l) irá melhorar o critério do problema de optimização relativo a esse AP face a qualquer elemento do mapa que registasse menos que -80 dBm. Esta situação é um contra-senso, porque tendo l registado ausência de sinal no AP i deveria piorar este critério relativamente a qualquer elemento do mapa que tivesse registado sinal, uma vez que a ausência de sinal significaria um valor infinitamente baixo no modelo log-distância.

Por este motivo desenvolveu-se um algoritmo que constrói um mapa de cobertura

de sinal com a mesma informação que um mapa Λ nas componentes das leituras diferentes de 0 e com o valor -100 nas componentes das leituras de Λ com o valor 0. Como na prática os valores da intensidade de sinal registados em Λ são superiores a -90 dBm, qualquer elemento do mapa que tenha registado sinal irá melhorar o critério correspondente a esse AP . Portanto, será uma tradução mais fiável do impacto de uma leitura com ausência de sinal num mapa de cobertura de sinal do que o valor 0. Para facilitar a notação, dado um mapa de cobertura de sinal Λ que represente ausência de sinal por 0, o mapa obtido por este processo será denotado por Λ^{100} .

Este mapa foi construído com o fim de testar o desempenho de alguns dos algoritmos desenvolvidos nos mapas Λ e Λ^{100} . Deste modo, pretende-se verificar se o tratamento mais adequado da ausência de sinal tem impacto no rendimento das abordagens de localização.

3.7. Média Móvel

Quando se testou os algoritmos multicritério em tempo real, verificou-se a instabilidade da localização estimada enquanto o dispositivo a localizar se mantinha no mesmo local. Esta situação ocorre devido às flutuações dos sinais registados nos APs .

Uma ideia para contornar este problema será considerar a média móvel de período p da sucessão de leituras recolhidas, que é descrita do seguinte modo:

Definição 9 (Média Móvel Simples). *Seja A_n , $i = 1, \dots, n$ as sucessivas observações de uma variável aleatória A e $p \in \{1, \dots, n\}$.*

$$A \text{ média móvel de período } p \text{ é dada por } \frac{\sum_{i=n-(p-1)}^n A_i}{p}.$$

Implementou-se portanto um algoritmo que calculava a média móvel de período p . Assim no programa de localização que corre em tempo real, a leitura que se considera para localizar é a média das últimas p leituras registadas pelo dispositivo. Repare-se que a implementação da média móvel implica um atraso no reconhecimento da posição real. Esse atraso será proporcional ao período p considerado.

Na situação prática considerou-se uma média móvel de período 5. Como as leituras são feitas de 2 em 2 segundos (aproximadamente), espera-se um atraso igual a 10 segundos relativamente à posição actual.

Dada a exigência de 5 leituras, após a execução do programa de localização,

primeiro recebem-se 5 leituras do dispositivo, só depois o programa começa a estimar a localização a partir da média móvel.

O algoritmo que calcula as médias móveis guarda as 5 últimas leituras. Assim, cada vez que o programa recebe uma nova leitura do dispositivo elimina a leitura mais antiga que foi registada, continuando a guardar 5 leituras. Depois calcula a média componente a componente das componentes diferentes de 0 ou -100 dessas cinco leituras, isto é, dos *APs* que registaram sinal. Para exemplificar o cálculo dessa média suponha-se que as 5 leituras referentes a 3 *APs* são:

$$l_1 = (-80, -40, 0)$$

$$l_2 = (0, -40, 0)$$

$$l_3 = (-90, -40, 0)$$

$$l_4 = (0, -40, 0)$$

$$l_5 = (0, -45, 0)$$

A média dessas cinco leituras é $(-85, -41, 0)$. Repare-se que se uma leitura não registou sinal de um *AP* não se considera na média e que só se nas 5 leituras um *AP* não estava visível a sua componente terá o valor 0.

Capítulo 4

Testes Computacionais

4.1. Descrição dos Testes

Os dados para os testes dos algoritmos foram obtidos em cenários reais, um lar de idosos na Trofa em que a *OnCaring* permitiu a obtenção de resultados e a sala 4.4 do Departamento de Matemática da Universidade de Coimbra.

Elaboraram-se dois mapas de cobertura de sinal do lar de idosos na Trofa para testes do filtro de Pareto a nível de multi-piso. O mapa referente ao primeiro piso também foi usado para analisar o desempenho dos diversos algoritmos a nível de piso. O mapa da sala 4.4 do Departamento de Matemática da Universidade de Coimbra, foi utilizado para demonstração dos algoritmos de Pareto em tempo real.

Na elaboração dos mapas usou-se o mesmo método de recolha. Este método consiste em recolher várias leituras na mesma posição em tempos consecutivos, com a particularidade de obrigar a disposição das recolhas das várias leituras formarem uma grelha, com espaçamento uniforme de aproximadamente um metro. Aplica-se este método com o objectivo de obter um novo mapa mais robusto e regular em que a leitura associada a uma posição corresponde à média das leituras com essa posição no mapa original em que se fez a recolha.

O mapa de cobertura de sinal do primeiro piso do lar de idosos da Trofa recolhido, contém 3631 leituras referentes a 10 *APs* com cerca de 10 leituras recolhidas por posição. O novo mapa com as médias das leituras obtido a partir deste contém 362 elementos e é o mapa usado pelos algoritmos para localizar neste piso. Apresentam-se as posições, em píxeis, das suas leituras na figura 4.1.

Com o propósito de estudar o erro destes algoritmos recolheram-se 569 leituras na região dos dados do mapa de sinal em que a posição é conhecida. Deste modo, testar-se-á os resultados de cada algoritmo na localização destas leituras.

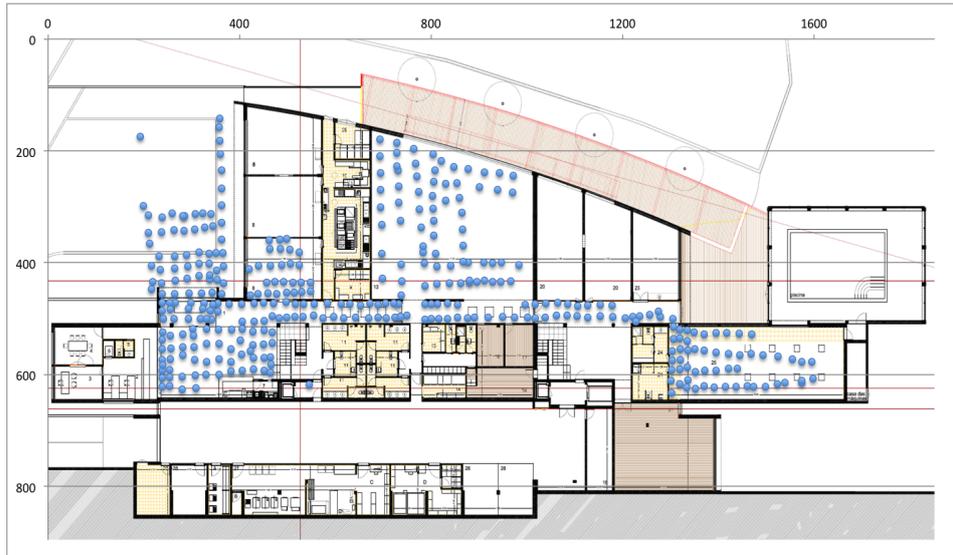


Figura 4.1: Posições do mapa de cobertura de sinal do piso 1 da Trofa

A unidade das posições no mapa de cobertura de sinal é o píxel. Para que o erro fosse dado em metros verificou-se a relação entre píxeis e metros neste mapa. Este cálculo indicou que 13 píxeis na imagem do piso correspondiam a 0.8 metros.

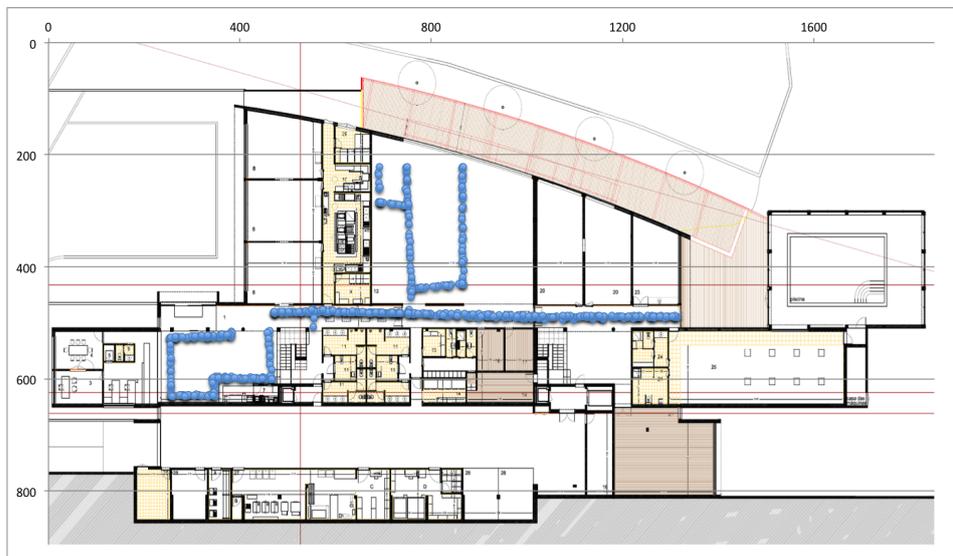


Figura 4.2: Leituras recolhidas no piso 1 para teste

4.2. Análise do Erro

A comparação dos algoritmos será realizada com a análise da eficiência e precisão das estimativas obtidas. A análise do erro consiste em descrever o erro da estratégia de cada algoritmo que apresenta o menor erro médio relativo às 569 leituras recolhidas

no primeiro piso da Trofa, a considerarem no máximo 10 *APs*.

O erro é descrito pelo seu valor mínimo, médio e máximo, bem como pelos quartis e desvio-padrão. Também se apresentam algumas descrições gráficas, como o *boxplot*, o histograma e a curva de frequências relativas acumulada. Estes elementos permitem identificar *outliers* e mostrar como o erro se encontra distribuído em termos da sua frequência por classe.

Em primeiro lugar estuda-se o erro dos algoritmos com uma abordagem multicritério e só posteriormente se estuda o algoritmo dos k vizinhos mais próximos.

4.2.1. Erro do Filtro de Pareto

Algoritmo FPDR

Estuda-se o erro do algoritmo que considera os k *APs* da leitura a localizar com maior intensidade de sinal e a relação de dominância de grau n . Note-se que este algoritmo quando considera todos os *APs* e a relação de dominância de grau 1 coincide com a formulação do problema (2.1) e, portanto, a primeira formulação multicritério do problema de localização *Indoor* também é analisada.

Apresenta-se um *contour plot* do erro médio em função do número de critérios melhorados e do número dos *APs* com melhor intensidade de sinal considerados na leitura a localizar.

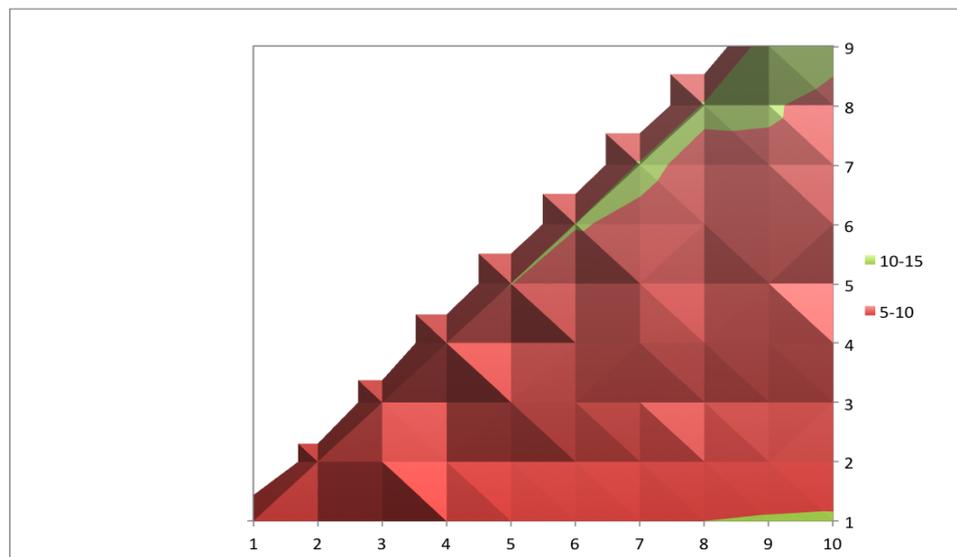


Figura 4.3: Erro do algoritmo **FPDR**

O eixo vertical representa o número de critérios melhorados e o eixo horizontal o número de *APs* com melhor intensidade de sinal considerados.

Em suma, a estratégia na qual se obteve menor erro médio, de 5.44 metros,

consiste em considerar a relação de dominância de grau 3 e os critérios dos 9 *APs* com maior intensidade de sinal na leitura.

Apresentam-se alguns parâmetros do erro em metros, dos quais se destaca um elevado erro máximo, de 31.05 metros.

Mínimo	$Q_{1/4}$	Mediana	$Q_{3/4}$	Máximo	Média	Desvio-Padrão
0.09	2.84	4.59	6.99	31.05	5.44	3.85

O *boxplot* da figura 4.4 evidencia uma assimetria positiva e ilustra a existência de 23 *outliers*, que correspondem aos valores de erro superiores a 13.2 metros.

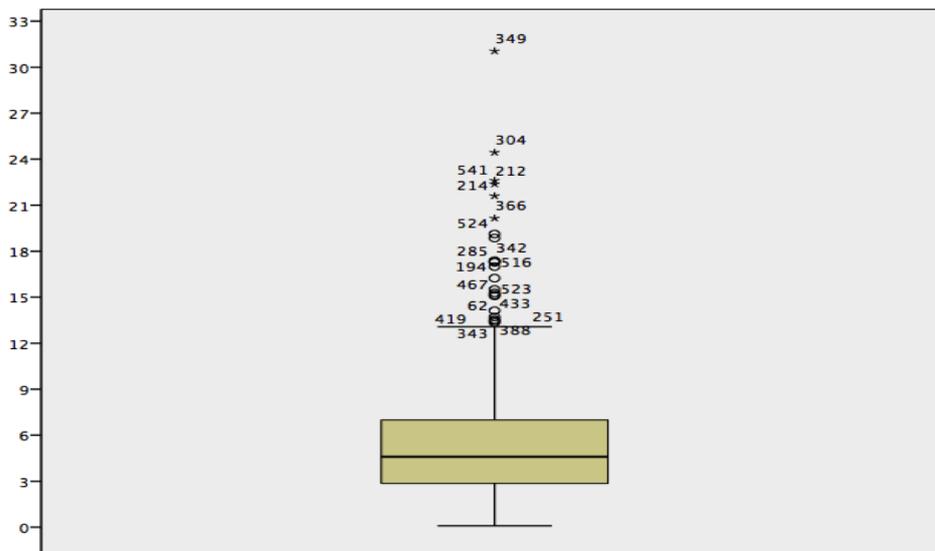


Figura 4.4: *Boxplot* do erro para melhor estratégia do algoritmo **FPDR**

Uma das principais razões encontradas para justificar os valores extremos das leituras 349, 304 e 541, foi que a maior parte das soluções de Pareto melhoram principalmente nos 4 ou 5 critérios de *APs* que a leitura não registou sinal e nos restantes critérios, em geral, diferem em cerca de 10 decíbel.

Por fim, apresenta-se um histograma na figura 4.5 e um gráfico da frequência relativa acumulada do erro deste algoritmo na figura 4.6.

Verifica-se, neste algoritmo, que a classe do erro entre os 4 e 5 metros apresenta maior frequência, 82 das 569 observações. Além disso, 77 observações obtiveram um erro entre os 3 e 4 metros. Estes valores indicam que 27.9% das observações obtiveram um erro entre os 3 e 5 metros. Além disso, o gráfico das frequências relativas, mostra que o algoritmo estimou cerca de 25% das observações com um erro inferior a 3 metros. Isto indica que este algoritmo embora obtenha bons resultados é prejudicado por observações completamente díspares, por exemplo apresentar um

erro de 31 metros. Também se conclui que pode estimar uma boa localização grosseira em cerca de 79.4% dos casos, que corresponde à percentagem de observações com erro inferior a 7.5 metros.

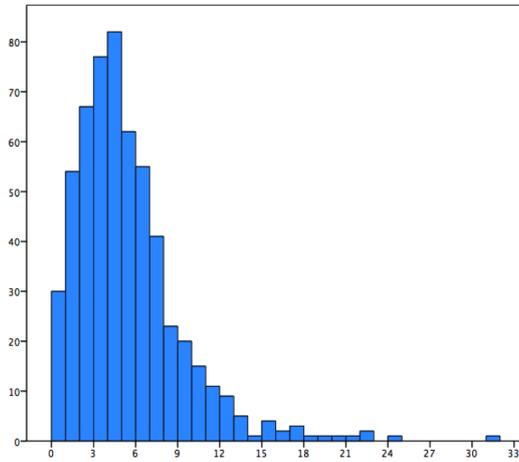


Figura 4.5: Histograma do erro

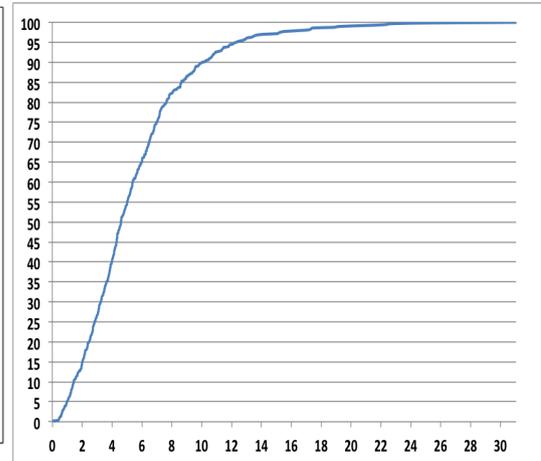


Figura 4.6: Frequência relativa acumulada

Algoritmo FPAPV

Analisa-se o erro do filtro de Pareto quando se considera somente os critérios de *APs* visíveis na leitura a localizar e a relação de dominância de grau n .

Este foi o algoritmo escolhido para analisar a diferença entre os resultados obtidos quando se calcula a estimativa da localização através da média ponderada das soluções de Pareto todas com o mesmo peso, com as estimativas obtidas através do cálculo da média ponderada MP das várias soluções.

Em primeiro lugar, apresenta-se a figura 4.7 que descreve os vários parâmetros do erro em função do número de critérios melhorados pelo algoritmo **FPAPV** quando atribui o mesmo peso a todas as soluções.

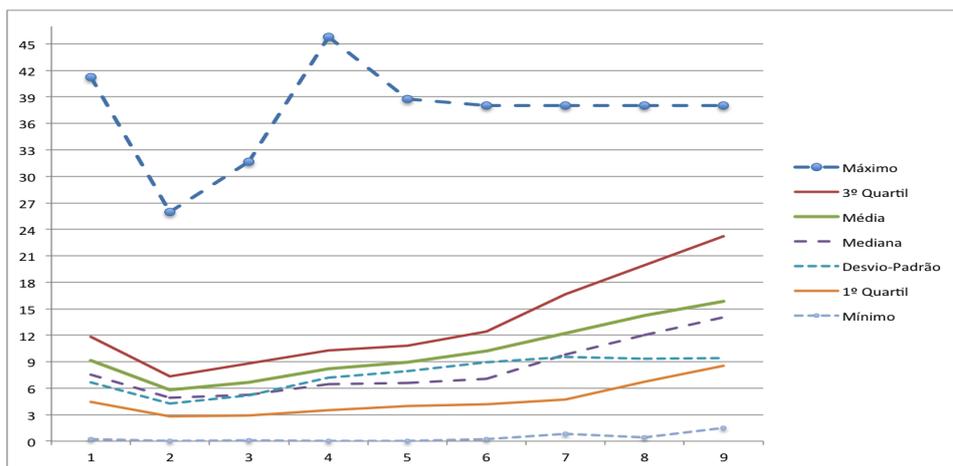


Figura 4.7: Erro do algoritmo **FPAPV** com centro

O menor erro médio obteve-se ao considerar-se a relação de dominância de grau 2, com um valor de 5.81 metros.

Finalmente, apresenta-se o gráfico com os parâmetros do erro em função da relação de dominância considerada por este algoritmo quando este estima a posição da leitura através da média ponderada MP.

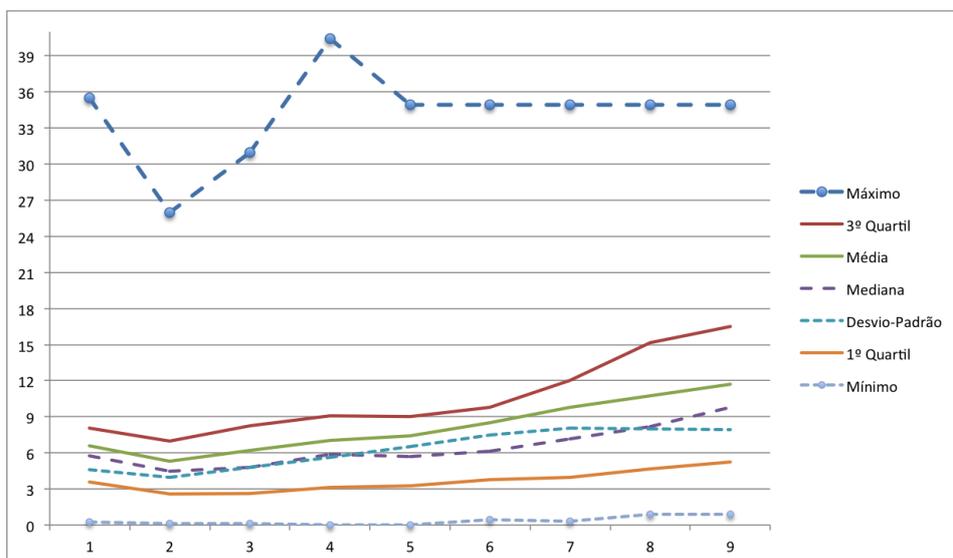


Figura 4.8: Erro do algoritmo **FPAPV** com média ponderada MP

A melhor estratégia consiste em considerar a relação de dominância de grau 2, em que o erro médio é de 5.298 metros. A figura 4.8 mostra que para cada estratégia os parâmetros do erro são, em geral, inferiores relativamente ao mesmo algoritmo quando atribui o mesmo peso a cada solução de Pareto. Este facto parece indicar que afectar cada solução de Pareto com um peso inversamente proporcional à sua

distância com a leitura l^* melhora a precisão do resultado.

Possivelmente obtiveram-se melhores resultados com a relação de dominância de grau 2 do que as restantes estratégias porque ao impôr-se melhorar mais do que 2 critérios é uma condição muito forte já que, em média, uma leitura regista 7 *APs*. Deste modo, devem existir mais elementos não dominados e por isso, em geral, é provável que hajam soluções mais afastadas da posição da leitura e, portanto, a estimativa seja pior.

Enumeram-se os vários parâmetros do erro para as melhores estratégias na seguinte tabela.

	Mínimo	$Q_{1/4}$	Mediana	$Q_{3/4}$	Máximo	Média	Desvio-Padrão
Centro	0.04	2.81	4.9	7.35	25.96	5.81	4.27
MP	0.13	2.59	4.47	6.98	25.96	5.298	3.95

Novamente, o erro máximo obtido é acentuado, de 25.96 metros. Contudo, o desvio-padrão do erro quando se considera a média ponderada MP é relativamente inferior, cerca de 32 centímetros.

A figura 4.9 evidencia em ambos casos uma assimetria positiva. Além disso, ilustra a existência de 33 *outliers* quando é considerado o centro, que correspondem aos valores de erro superior a 14.147 metros. Por outro lado, o *boxplot* da média ponderada MP evidencia a existência de 25 *outliers*, que correspondem aos valores de erro superior a 13.57 metros.

Quando se examina cada *boxplot* da figura 4.9 verifica-se que muitos dos *outliers* quando se atribuiu o mesmo peso a cada solução também são *outliers* no caso de se considerar a média ponderada, por exemplo, as leituras 393, 289, 357, 307.

As leituras 357 e 307 registaram apenas 3 *APs* visíveis e em cada uma destas leituras o algoritmo obteve uma única solução de Pareto bastante afastada da posição da leitura.

Ao analisarem-se os *outliers* verificou-se que registavam no máximo 7 *APs*. Neste caso, a relação de dominância de grau 2 possivelmente é uma exigência muito forte. Na medida de explorar esta possibilidade testou-se o algoritmo **FPAPV** nos *outliers* com a relação de dominância de grau 1. O erro máximo destas leituras reduziu-se para 18 metros e a maior parte do erro destas leituras reduziu-se para menos de 9 metros, com o erro de 4 leituras inferior a 3.6 metros.

Segundo este raciocínio, se fosse possível captar as propriedades dos *outliers*,

poder-se-ia procurar uma estratégia melhor para esses casos e obter melhores estimativas.

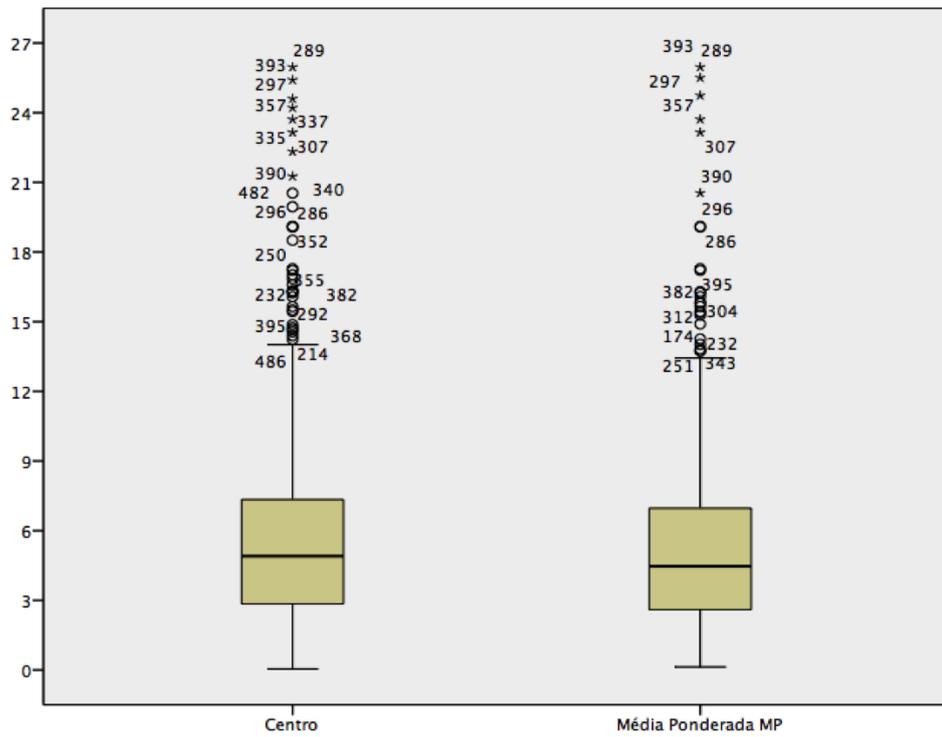


Figura 4.9: Boxplot do erro do algoritmo **FPAPV**

Apresenta-se o histograma e a frequência relativa acumulada do erro deste algoritmo quando considera o centro das soluções de Pareto como estimativa da localização.

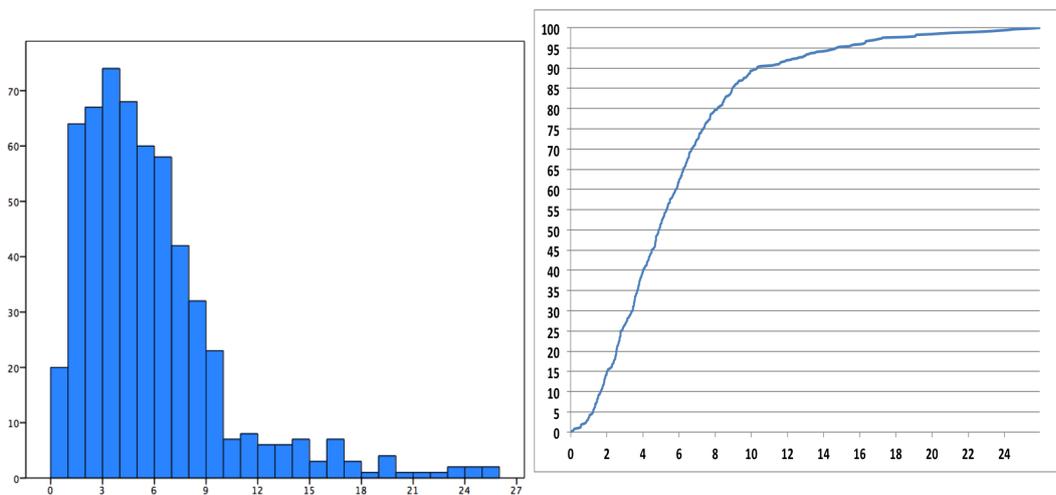


Figura 4.10: Histograma do erro

Figura 4.11: Frequência relativa acumulada

O histograma indica que a classe do erro entre os 3 e 4 metros tem maior frequência, além disso, o gráfico com a frequência relativa acumulada mostra que cerca de 26.5% das posições estimadas das leituras têm um erro inferior a 3 metros. Por fim, constata-se que 76% das observações têm um erro inferior a 7.5 metros.

Para terminar a descrição dos testes computacionais do algoritmo **FPAPV** estuda-se a distribuição do erro quando este estima a localização com a média ponderada MP.

A partir das figuras 4.12 e 4.13 inerentes a este algoritmo com a média ponderada MP, conclui-se que o erro entre os 3 e 4 metros ocorreu com maior frequência, em 85 observações, cerca de 15% das posições estimadas. Verifica-se também que 44.3% das observações têm um erro inferior a 4 metros. Por fim, constata-se que cerca de 80% das posições estimadas têm um erro inferior a 7.5 metros.

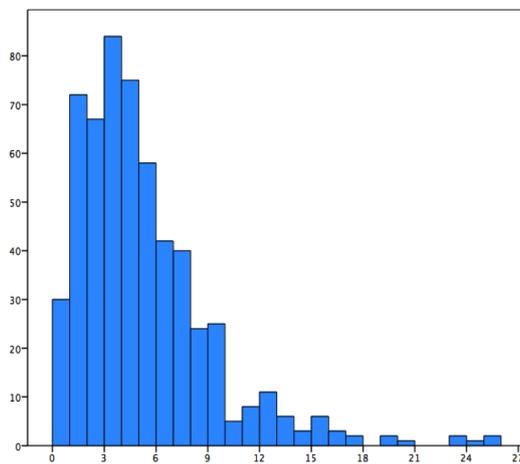


Figura 4.12: Histograma do erro

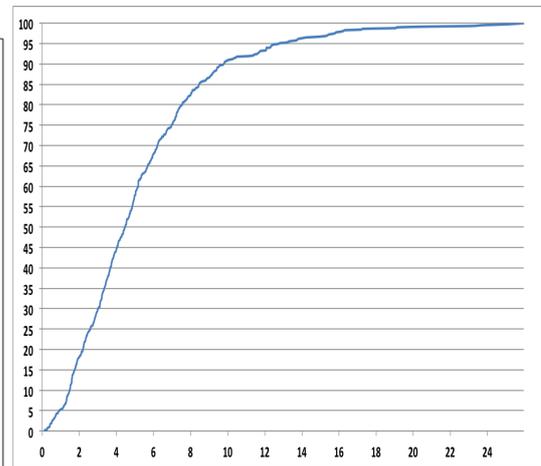


Figura 4.13: Frequência relativa acumulada

4.2.2. k Vizinhos Mais Próximos

Analisa-se o erro dos k vizinhos mais próximos comparativamente a duas métricas. A primeira métrica que é analisada trata-se da euclidiana normalizada, a segunda, a de *Mahalanobis*. Além destes também é analisado o algoritmo **KVAPV**.

A figura 4.14 representa o erro médio dos vários algoritmos em função do número de vizinhos considerados.

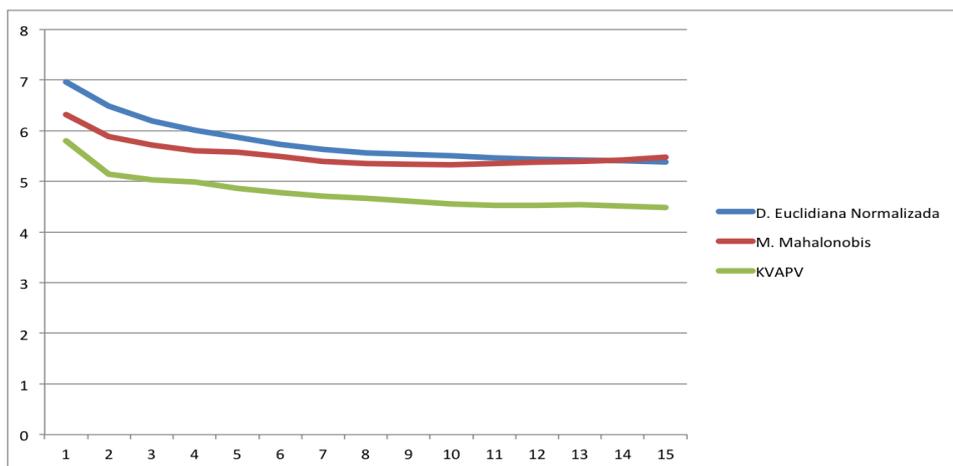


Figura 4.14: Erro dos k vizinhos mais próximos

Em todos os algoritmos o erro médio ao considerarem-se entre 8 a 14 vizinhos é muito próximo, deste modo não compensa computacionalmente escolher mais vizinhos. Com base na figura 4.14 e no raciocínio anterior, a melhor estratégia consiste em escolher 8 vizinhos em cada algoritmo pois obtém-se um erro médio inferior às estratégias que consideram menos que 8 vizinhos.

Descreve-se os parâmetros do erro na seguinte tabela. Todos os algoritmos obtiveram um erro máximo acentuado, no entanto o algoritmo **KVAPV** apresenta valores inferiores praticamente em todos os parâmetros.

	Mínimo	$Q_{1/4}$	Mediana	$Q_{3/4}$	Máximo	Média	Desvio-Padrão
KVDEN	0.06	2.56	4.68	7.239	25.62	5.56	4.44
KVMM	0.06	2.35	4.82	7.19	34.38	5.35	3.91
KVAPV	0.08	2.09	3.86	6.58	22.32	4.66	3.35

A figura 4.15 indica uma assimetria positiva no erro dos 3 algoritmos. No algoritmo que considera a distância euclidiana normalizada (**KVDEN**) as 26 leituras com erro superior a 14.2575 metros são *outliers*, e no que considera a métrica de *Mahalanobis* (**KVMM**), são *outliers* as 18 leituras com erro superior a 14.45 metros. Por último, no algoritmo **KVAPV** é evidente a existência de 10 *outliers* que correspondem às leituras com erro superior a 13.32 metros.

A observação 214 é um *outlier* nos 3 algoritmos. Analisando esta leitura, verifica-se que os 8 elementos mais próximos, em termos de sinal, estão a uma distância inferior a 13 decíbel, muito próximos. No entanto as posições destes elementos encontram-se a mais de 8 metros da posição desta leitura. Isto sucedeu-se possivelmente por ter

ocorrido um pico de sinal em vários APs .

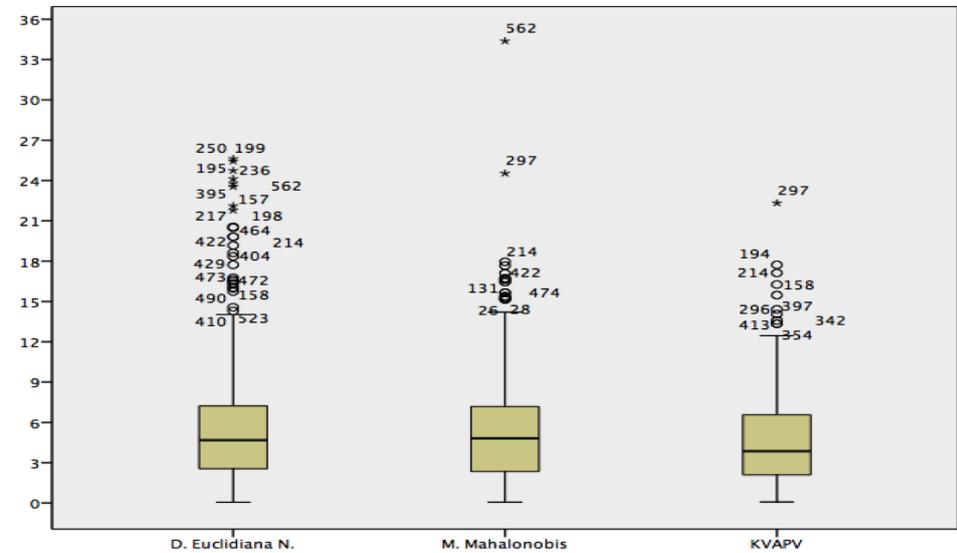


Figura 4.15: *Boxplot* do erro dos k vizinhos mais próximos

Para concluir, apresenta-se um histograma e um gráfico das frequências relativas acumuladas para o algoritmo **KVAPV**.

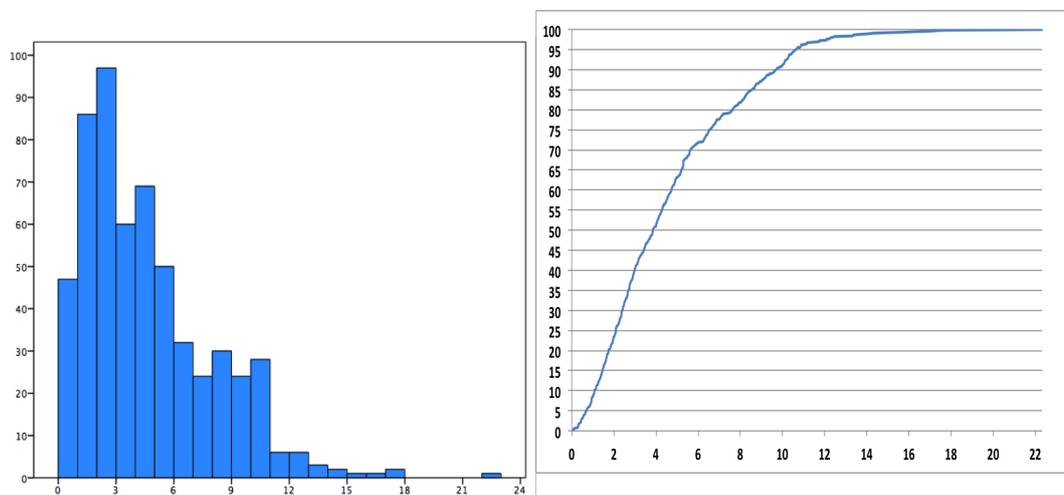


Figura 4.16: Histograma do erro

Figura 4.17: Frequência relativa acumulada

A classe do erro entre os 2 e 3 metros apresenta maior frequência, 97 das 569 observações totais, aproximadamente 17% das leituras. Além disso, cerca de 31% das observações têm um erro entre 1 a 3 metros. Para concluir, constata-se também que 79% das observações têm um erro inferior a 7.5 metros.

4.2.3. k Vizinhos de Pareto Mais Próximos

Para terminar a análise dos algoritmos a nível de piso estuda-se o erro do algoritmo dos k vizinhos de Pareto mais próximos.

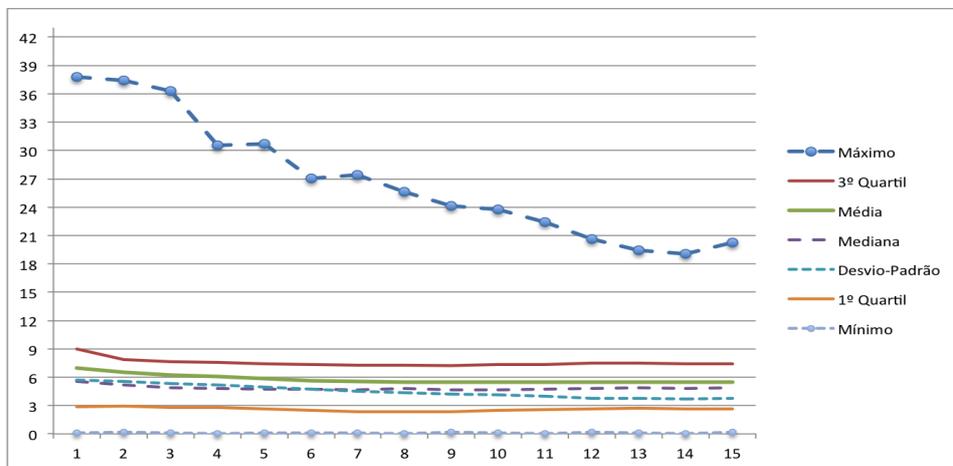


Figura 4.18: Erro dos k vizinhos de Pareto mais próximos

A melhor estratégia consiste em considerar os 9 vizinhos mais próximos, na qual se obteve um erro médio de 5.48 metros, ligeiramente inferior aos k vizinhos com distância euclidiana normalizada.

Conclui-se que o filtro de Pareto neste *data set* elimina alguma informação geométrica das leituras relativamente à métrica euclidiana.

Veja-se o seguinte exemplo, suponha-se que os 9 elementos mais próximos em termos de sinal também eram soluções de Pareto, assim o resultado deste algoritmo seria igual ao resultado do algoritmo dos k vizinhos mais próximos para $k = 9$. Contudo como se obtêm valores de erro diferentes entre os algoritmos, isto é, resultados diferentes em alguns casos, conclui-se que nem todos os 9 vizinhos mais próximos eram soluções de Pareto.

Apresentam-se alguns parâmetros do erro para a melhor estratégia. Repare-se que os vários parâmetros são ligeiramente inferiores aos do algoritmo dos k vizinhos mais próximos com a distância euclidiana normalizada, principalmente o desvio-padrão.

Mínimo	$Q_{1/4}$	Mediana	$Q_{3/4}$	Máximo	Média	Desvio-Padrão
0.17	2.37	4.67	7.22	24.14	5.48	4.23

O *boxplot* da figura 4.19 apresenta uma assimetria positiva e evidencia a existência de 22 *outliers* que correspondem às leituras com erro superior a 14.49 metros.

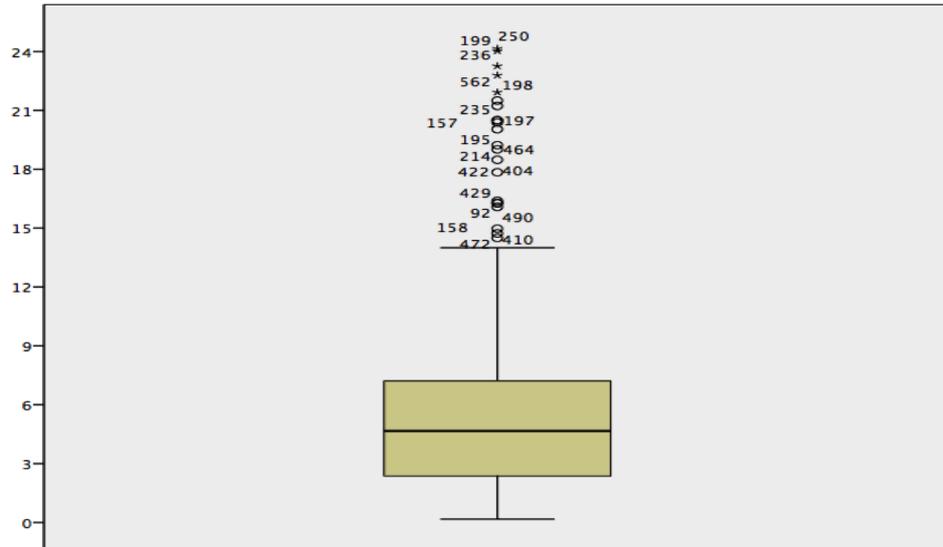


Figura 4.19: *Boxplot* do erro dos k vizinhos de Pareto mais próximos

O erro das leituras 562, 250, 199 e 236 no algoritmo dos k vizinhos mais próximos com a distância euclidiana normalizada também foram valores extremos. Numa análise a estes *outliers*, constata-se que a distância mínima, em termos de sinal, das leituras das soluções de Pareto é cerca de 100 decíbel. Ou seja, nenhuma das soluções de Pareto se aproximava em termos de sinal da leitura dada. Por outro lado, também se constatou o mesmo com o algoritmo dos k vizinhos mais próximos.

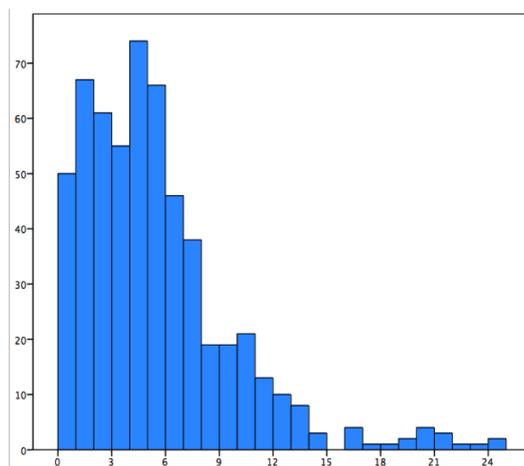


Figura 4.20: Histograma do erro

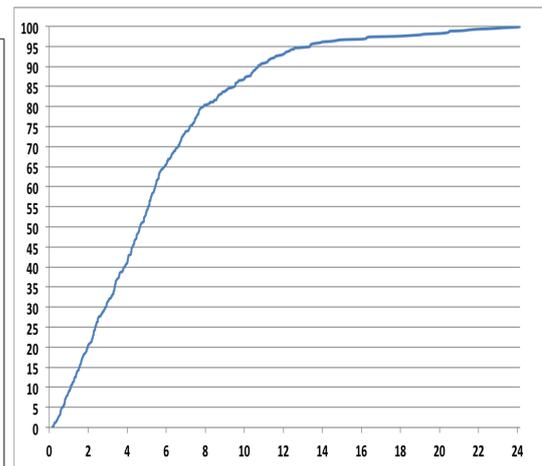


Figura 4.21: Frequência relativa acumulada

Em suma, a classe do erro entre os 4 e 5 metros apresenta maior frequência, 72 das 569 observações totais. Além disso, cerca de 41% das observações têm um erro inferior a 4 metros e 77% apresentam um erro inferior a 7.5 metros. Isto indica

que o algoritmo pode estimar uma boa localização grosseira, com cerca de 41% das posições estimadas com um erro inferior a 4 metros.

4.3. Resultados Multi-Piso

Testou-se o algoritmo **FPAPV** estendido a multi-piso considerando a relação de dominância de grau 2. Com o intuito de comparar com algoritmos da literatura testaram-se os mesmos dados com o algoritmo dos k vizinhos mais próximos estendido a multi-piso com um processo análogo ao filtro de Pareto.

Para teste consideraram-se o primeiro e o segundo piso da Trofa, com 569 leituras para teste localizadas no primeiro piso e 632 leituras de teste localizadas no segundo piso.

As posições das leituras do mapa de cobertura de sinal do primeiro piso foram representadas anteriormente, por isso apresentam-se simplesmente as 203 posições do mapa de cobertura de sinal do segundo piso do lar de idosos da Trofa relativo a 12 APs.

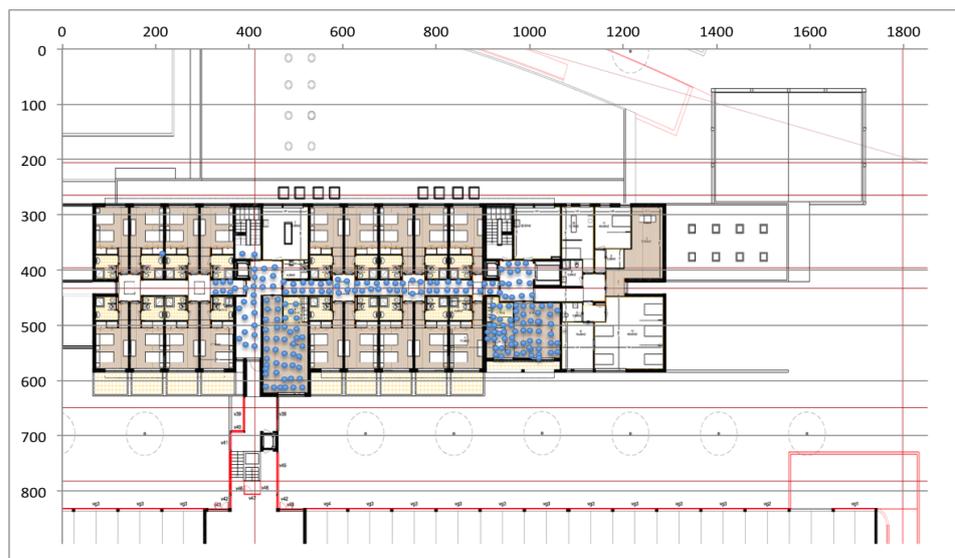


Figura 4.22: Posições do mapa de cobertura de sinal do piso 2 da Trofa

Além disso, enquadram-se as posições das 632 leituras recolhidas para teste no segundo piso na planta desse piso. Note-se que em ambas as figuras a unidade da posição é dada em píxel.

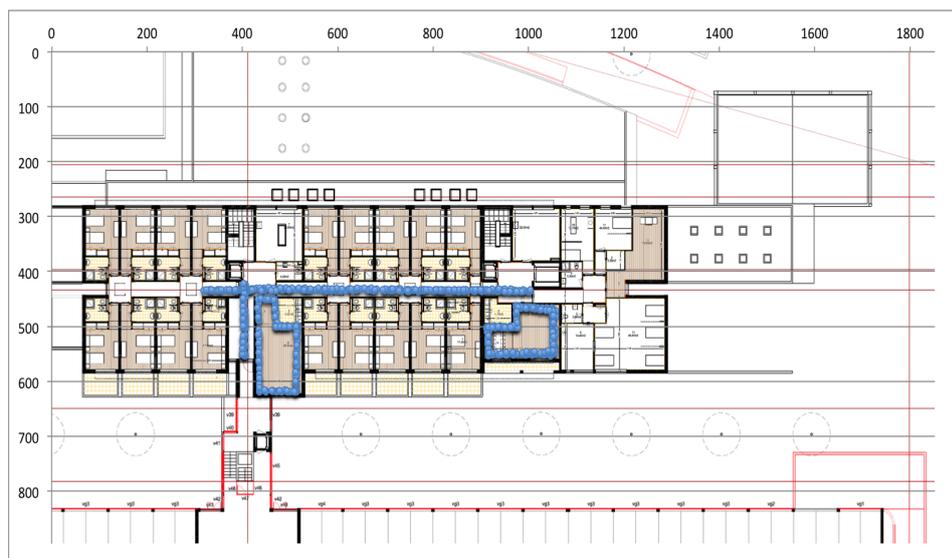


Figura 4.23: Posições das leituras para teste no piso 2 da Trofa

Finalmente, descrevem-se os resultados obtidos relativos à identificação correcta do piso através do filtro de Pareto na seguinte tabela, em que “falhou por empate” significa ter obtido o mesmo número de soluções em cada piso.

	Acertou	Falhou	Falhou por Empate
Piso 1	549	12	8
Piso 2	632	0	0

Em suma, o filtro de Pareto Multi-piso relativamente às 1201 leituras teve uma taxa de sucesso de 98.3%. Por outro lado, o algoritmo dos k vizinhos mais próximos teve uma taxa de sucesso de 98.7%.

Acrescenta-se uma breve nota. O filtro de Pareto Multi-piso poderia ser melhorado se houvesse maneira de desempatar os 8 casos apresentados com um critério coerente. Um exemplo de um possível critério seria determinar o piso com a solução de Pareto mais próxima da leitura em termos de sinal.

4.4. Conclusões

Numa análise sumária ao erro dos vários algoritmos para as melhores estratégias, tiram-se as seguintes conclusões.

Todos os algoritmos apresentam uma variação enorme entre o erro mínimo e máximo. Os dados sugerem que o erro apresente uma elevada dispersão relativamente à média para qualquer algoritmo tendo em conta o desvio-padrão obtido pelos vários algoritmos.

Com o fim de equiparar os dados do erro dos diversos algoritmos multicritério constrói-se a seguinte tabela com a mediana do erro, o erro médio e o desvio-padrão, em metros, para as melhores estratégias de cada um.

	Mediana	Média	Desvio-Padrão
FPDR	4.59	5.44	3.85
FPAPV - Centro	4.9	5.81	4.27
FPAPV - MP	4.47	5.298	3.95
KVP	4.67	5.48	4.23

Relativamente a estes dados parece credível aceitar que o algoritmo **FPAPV** com média ponderada parece obter as melhores estimativas da posição. Todavia, deve-se recorrer a métodos estatísticos para verificar se essa hipótese é válida.

Seja μ a média da população que representa o erro do algoritmo **FPAPV** com o centro e $\bar{\mu}$ a média do erro deste algoritmo com a média ponderada MP. Utilizando o *software SPSS* calculou-se o p-valor do seguinte teste de hipótese.

$$H_0 : \mu = \bar{\mu} \text{ vs } H_1 : \mu > \bar{\mu}.$$

O p-valor deste teste foi 0, pelo que se rejeita a hipótese H_0 . Assim, a evidência contra H_0 é significativa relativamente a qualquer valor de significância.

Deste modo, em princípio, o algoritmo **FPAPV** com a média ponderada MP apresentará melhores resultados do que com o centro.

Por outro lado, repetiu-se o teste mas com μ a representar a média da população do erro dos k vizinhos de Pareto. Neste caso, o p-valor foi de 0.2045, pelo que se aceita a hipótese H_0 nos níveis de significância usuais, ou seja, de 0.01, 0.05, 0.1 e 0.2. Desta forma, não há indícios de que o comportamento do algoritmo **FPAPV** com média ponderada seja melhor do que os k vizinhos de Pareto mais próximos.

Repetiu-se também o teste com μ a representar a média da população do erro do algoritmo **FPDR**. Obteve-se um p-valor de 0.208, superior aos níveis de significância usuais. Por isso, também não se pode inferir que as médias são diferentes, ou seja, aceita-se que as médias do algoritmo **FPDR** e do algoritmo **FPAPV** com média ponderada MP não sejam diferentes, ao nível de significância 0.1 e 0.2.

Verifica-se também que os algoritmos multicritério propostos permitem estimar uma localização grosseira em cerca 80% dos casos e são boas aproximações iniciais para que outros algoritmos possam refinar essa estimativa.

Para terminar, conclui-se que o algoritmo multicritério multi-piso é suficiente-

mente bom para determinar o piso em cerca de 97.3% dos casos, isto é apresenta uma margem de erro bastante aceitável.

Capítulo 5

Trabalho Futuro

Como trabalho futuro pretende-se desenvolver uma versão do Filtro de Pareto que como *data set* não considera um conjunto discreto de pontos, como o mapa de cobertura de sinal, mas sim um modelo de propagação de sinal ajustado ao edifício.

Este estudo já foi iniciado, mas está ainda numa fase inicial. Descreve-se assim a ideia para resolver este problema.

Se um piso de um edifício for delimitado por uma região $\bar{R} \subset \mathbb{R}^2$ com m APs calcula-se o seu modelo contínuo de propagação de sinal P_m dado por,

$$\begin{aligned} P_m : \quad R &\longrightarrow \mathbb{R}^m \\ (x, y) &\longmapsto (P_{m1}(x, y), \dots, P_{mm}(x, y)) \end{aligned}$$

Para cada leitura $l^* = (l_1, \dots, l_m)$, pretende-se resolver o problema

$$\min_{(x,y) \in \bar{R}} G_{l^*, P_m}(x, y),$$

onde $G_{l^*, P_m}^i(x, y) = (l_i - P_{mi}(x, y))^2$.

Este problema de optimização implica o cálculo das soluções de Pareto. Como o conjunto de soluções de Pareto pode ser infinito, este problema passa também por estudar qual o seu subconjunto finito que deve ser escolhido para estimar a localização.

Para calcular soluções de Pareto introduz-se um teorema dado em [Ehr05] que permite resolver um problema de minimização de uma função real para determinar soluções de Pareto.

Teorema 5. *Seja $F : X \longrightarrow \mathbb{R}^m$ a função objectivo de um problema de optimização multicritério com conjunto admissível X e $\lambda \in (\mathbb{R}^+)^m$.*

A solução \bar{x} de $\min_{x \in X} \langle \lambda, F(x) \rangle$ é uma solução óptima de Pareto.

Demonstração. Suponha-se que \bar{x} não era solução óptima de Pareto, ou seja, que existe $x' \in X$ com $F(x') \leq F(\bar{x})$ e para o qual $\exists i, 1 \leq i \leq m, F_i(x') < F_i(\bar{x})$. Assim

ter-se-ia para $1 \leq j \neq i \leq m$, $\lambda_j F_j(x') \leq \lambda_j F_j(\bar{x})$ e $\lambda_i F_i(x') < \lambda_i F_i(\bar{x})$. Isto implica

$$\langle \lambda, F(x') \rangle = \sum_{h=1}^m \lambda_h F_h(x') < \sum_{h=1}^m \lambda_h F_h(\bar{x}) = \langle \lambda, F(\bar{x}) \rangle .$$

O que contradiz a hipótese de \bar{x} ser o mínimo. □

Deste modo procurar-se-ia determinar $\lambda^1, \lambda^2, \dots, \lambda^k \in (\mathbb{R}^+)^m$ de forma a se obter boas soluções de Pareto. Isto consistiria em resolver k problemas de minimização,

$$\min_{(x,y) \in R} \langle \lambda_i, G_{l^*, P_m}(x, y) \rangle \quad (5.1)$$

Cada um dos quais se pode resolver com o método da descida mais rápida que é descrito do seguinte modo (considerou-se este método como uma abordagem inicial, porque o gradiente desta função é simples de calcular. Futuramente pretende-se estudar métodos mais eficientes e adequados para resolver este problema):

Definição 10 (Método da Descida Mais Rápida). *Seja $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável em X com gradiente ∇f e x_0 uma aproximação inicial para um mínimo de f .*

As iterações deste método são, para $n \in \mathbb{N}$, dadas por $x_{n+1} = x_n - \gamma_n \nabla f(x_n)$, com γ_n a satisfazer $f(x_n - \gamma_n \nabla f(x_n)) < f(x_n)$.

Para concluir, resume-se a ideia com que se pretende desenvolver o filtro de Pareto contínuo para uma leitura l^* e considerando um modelo P_m relativamente a m APs.

1. Determinar $k \in \mathbb{N}$ e $\lambda^1, \dots, \lambda^k \in \mathbb{R}^m$ de forma adequada;
2. resolver os k problemas de minimização (5.1) determinando as soluções óptimas de Pareto x_1, \dots, x_k através do método da descida mais rápida;
3. Estimar a localização através de x_1, \dots, x_k .

Além da otimização multicritério em contínuo pretende-se também melhorar o algoritmo discreto, no sentido de encontrar um método de eliminar soluções de Pareto provavelmente muito afastadas da posição da leitura.

Por fim, pretende-se também encontrar um método para determinar melhor a localização final através de uma média ponderada mais ajustada e em conseguir relacionar estimativas de posições de leituras consecutivas, num cenário em tempo real, de forma a excluir algumas soluções de Pareto em posições impossíveis conhecida a distância mínima que é possível percorrer no tempo entre duas leituras com um algoritmo desenvolvido em [San14].

Siglas

Sigla	Significado
<i>WLAN</i>	<i>wireless local area network</i>
<i>AP</i>	access point
FP	filtro de Pareto
FPRN	filtro de Pareto com relação de dominância de grau n
FPAPD	filtro de Pareto aos <i>APs</i> com valor superior a d
FPKAP	filtro de Pareto aos k <i>APs</i> com maior intensidade de sinal
FPDR	filtro de Pareto com duas restrições
FPAPV	filtro de Pareto aos <i>APs</i> visíveis
KVP	algoritmo dos k vizinhos de Pareto mais próximos
KVDEN	k vizinhos mais próximos com distância euclidiana normalizada
KVMM	k vizinhos mais próximos com métrica de <i>Mahalonobis</i>
KVAPV	algoritmo dos k vizinhos mais próximos dos <i>APs</i> visíveis

Bibliografia

- [Bol13] Jason Nobre Bolito. Algoritmos de localização *Indoor*. Tese de Mestrado, Faculdade de Ciências da Universidade de Coimbra, Junho 2013.
- [BP00] Paramvir Bahl and Venkata N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *IEEE INFOCOM 2000*, volume 3, pages 775–784, 2000.
- [CK13] Patrick Croak and Young Kim. Site survey guidelines for wlan deployment, April 2013.
- [CS12] Solomon Chan and Gunho Sohn. Indoor localization using wi-fi based fingerprinting and trilateration techniques for lbs applications. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII, 2012.
- [Ehr05] Matthias Ehrgott. *Multicriteria Optimization*. Springer, 2nd edition, 2005.
- [HLL07] Pat Banerjee Hui Liu, Houshang Darabi and Jing Liu. Survey of wireless indoor positioning techniques and systems. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 37(6), 2007.
- [JYLS13] Hyunjae Park Joo-Yub Lee, Cheal-Hwan Yoon and Jungmin So. Analysis of location estimation algorithms for wifi fingerprint-based indoor localization. In *Advanced Researches on Software Technology*, volume 19, pages 89 – 92, 2013.
- [MA04] R.T. Marler and J.S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26:369–395, 2004.

- [Per14] Ana Pereira. Multilateração para localização indoor. Tese de Mestrado submetida para avaliação, Departamento de Matemática da Universidade de Coimbra, 2014.
- [San14] Cátia Santos. Caminhos mais curtos em localização em espaços fechados. Tese de Mestrado submetida para avaliação, Departamento de Matemática da Universidade de Coimbra, 2014.
- [Sch11] Herbert Schildt. *Java, The Complete Reference*. McGraw-Hill, 8th edition, 2011.