

UNIVERSIDADE FEDERAL DE SANTA CATARINA
UFSC-SC

Mateus Bruscato

Técnicas Básicas de *Machine Learning* aplicadas ao problema
inverso de Tomografia por Impedância Elétrica

TRABALHO DE CONCLUSÃO DE CURSO

SANTA CATARINA
2021

UNIVERSIDADE FEDERAL DE SANTA CATARINA
UFSC-SC

Mateus Bruscato

Técnicas Básicas de Machine Learning aplicadas ao problema
inverso de Tomografia por Impedância Elétrica

TRABALHO DE CONCLUSÃO DE CURSO

Trabalho de conclusão de curso apresentado
à Universidade Federal de Santa Catarina,
como requisito para conclusão do Curso de
Graduação em Matemática Licenciatura (no-
turno), sob a orientação do Prof. Fábio
Júnior Margotti.

SANTA CATARINA
2021

Resumo

A tomografia por impedância elétrica é uma ferramenta que permite obter informações do interior de um corpo de prova, através da aplicação de correntes elétricas em sua fronteira, e mensuração das respectivas voltagens produzidas também na fronteira. Essa ferramenta tem aplicação em diversas áreas. Conforme [Aguilar \(2009\)](#), uma aplicação é no reconhecimento de padrões de escoamento de tubos de extração de petróleo. Porém, para sua aplicação é necessário resolver um problema inverso que é mal posto no sentido de Hadamard. O presente trabalho utiliza técnicas de Aprendizagem de Máquina ou *Machine Learning* para reconstruir uma característica da solução do problema inverso de tomografia por impedância elétrica, para classificar o padrão de escoamento bifásico e calcular a fração volumétrica de óleo dentro do tubo de extração. Apresentaremos os resultados dos testes realizados com as técnicas de regressão linear e regressão logística, a partir de dados simulados de padrões de escoamento.

Palavras chave: Tomografia por Impedância Elétrica, Problemas Inversos, Machine Learning.

Abstract

The Electrical Impedance Tomography (EIT) is a tool that allows obtaining information from the interior of a body, through the application of electrical currents at its border and measurement of the respective voltages produced at the border as well. This tool has application in several areas. According to [Aguilar \(2009\)](#), one application is in the recognition of flow patterns in oil extraction pipes. However, for its application, it is necessary to solve an inverse problem that is misplaced in Hadamard's sense. The present document uses *Machine Learning* techniques to reconstruct a feature of the solution of the inverse electrical impedance tomography problem, to classify the two-phase flow pattern, and calculate the volumetric fraction of oil inside the extraction tube. We will present the results of tests carried out with the techniques of linear regression and logistic regression, from simulated data of flow patterns.

Keywords: Electrical Impedance Tomography, Inverse Problems, Machine Learning.

Sumário

Sumário	5
1 INTRODUÇÃO	7
2 CONCEITOS DA TEORIA CLÁSSICA DE PROBLEMAS INVERSOS	9
2.0.1 Propriedades da Pseudo-inversa em espaços de Hilbert	9
2.0.2 Teoria da regularização	11
2.0.2.1 Método Tikhonov	14
2.0.2.2 Método tipo gradiente (Landweber)	15
3 TOMOGRAFIA POR IMPEDÂNCIA ELÉTRICA	19
3.0.1 O Modelo Contínuo	19
3.0.2 O Modelo Completo de Eletrodos	22
4 CONCEITOS E TÉCNICAS BÁSICAS DE MACHINE LEARNING	25
4.0.0.1 Regressão Linear	26
4.0.0.2 Regressão Logística Simples	27
4.0.0.3 High Bias e Overfitting	28
4.0.1 <i>Cross Validation</i>	29
5 ESCOAMENTO BIFÁSICO	31
5.0.1 Padrão de escoamento	31
5.0.2 Fração Volumétrica	32
6 APLICAÇÃO	33
7 EXPERIMENTOS NUMÉRICOS	34
7.1 Determinação da Fração Volumétrica	37
7.1.1 Método de Landweber	38
7.1.2 Método de Tikhonov	41
7.1.2.1 Comparação entre Método de Landweber e Método de Tikhonov	43
7.2 Determinação do padrão de escoamento	44
7.3 Dados com ruído	49
7.3.1 Determinação de Fração Volumétrica	50
7.3.2 Determinação do padrão de escoamento	53
8 CONSIDERAÇÕES FINAIS	56

REFERÊNCIAS	57
APÊNDICE A – ESPAÇOS NORMADOS E ESPAÇOS DE HILBERT	58
APÊNDICE B – OPERADORES LINEARES E LIMITADOS E OPERADORES COMPACTOS	62

1 Introdução

Há situações na vida real que se assemelham ao funcionamento de uma função matemática. Situações, como por exemplo, o valor que iremos pagar comprando certa quantidade de pães. Porém, existem casos em que não conhecemos alguns dos elementos da função, por exemplo, podemos ter os valores da função (imagem da função), e mesmo assim não conhecermos a função explicitamente e até mesmo não sabermos de quais valores de entrada (valores do domínio) da função resultaram no valor de saída (imagem).

Para estudar problemas como esses, existe a teoria dos Problemas Inversos (PI). Para **TARANTOLA**, na teoria de PI, de maneira geral, para “prever” resultados de alguma mensuração precisa-se do modelo do sistema sobre investigação e uma teoria física ligando os parâmetros do modelo aos parâmetros a serem mensurados, e isto é o que se denomina, na teoria dos problemas inversos, de “problema direto”. Já um problema inverso consiste em usar os resultados de uma observação para determinar os valores dos parâmetros que caracterizam o sistema sobre investigação.

Um problema prático onde essa teoria é aplicada é na Tomografia por Impedância Elétrica (TIE), que é um método utilizado para obter informações do interior inacessível de um corpo de prova, onde são posicionados eletrodos no contorno desse corpo, que emitem correntes elétricas alternadas e que após a emissão, captam a diferença de potencial elétrico. Tal utilidade encontrou aplicação em diversas áreas, como: Medicina, Geofísica, Química, Engenharia, dentre outros (**AGUILAR, 2009**).

Entre os métodos para resolução de Problemas Inversos, temos os métodos da área de *Machine Learning*. Uma tradução para *Machine Learning*, seria Aprendizado de Máquina, e uma definição antiga dada por Arthur Samuel (1959), para tal área é: o campo de conhecimento que fornece ao computador a habilidade de aprender sem ser explicitamente programado.

Assim, o presente trabalho trata do estudo feito sobre tópicos acima para o teste da eficácia da regressão linear, regressão logística e do método de Tikhonov através da *Machine Learning* na resolução do problema inverso de TIE, mais especificamente na utilização da TIE, para reconstrução da condutividade elétrica do interior de tubos de escoamento de petróleo, e na classificação de padrões de escoamento no interior dos mesmos tubos. Os dados simulados utilizados para este estudo foram fornecidos pelo colaborador Eduardo Hafemann. O software utilizado para a realização dos experimentos numéricos foi o *MatLab R2013a*. A forma como os dados foram simulados é melhor detalhada na Seção 7.

O presente trabalho está organizado da seguinte maneira: na Seção 2 apresentamos alguns conceitos básicos da teoria de problemas inversos, como o que é um problema inverso e a definição de problema mal-posto segundo Hadamard. Na Subseção 2.1 ex-

ploramos algumas propriedades da pseudo-inversa em espaços de Hilbert, e mostramos a definição de problema mal-posto segundo Nashed. Já na Subseção 2.2 é explanada uma ideia geral de teoria de regularização para resolução de problemas inversos, e na sequência são apresentados os métodos de regularização de Tikhonov e Landweber.

Já na Seção 3 é apresentada uma ideia geral do que trata o problema inverso da tomografia por impedância elétrica, e nas subseções seguintes descrevemos matematicamente dois modelos da TIE: o modelo contínuo e o modelo completo de eletrodos.

Na sequência, explanamos na Seção 4 a definição de *Machine Learning* e de forma geral como são divididas as classes de algoritmos. Seguindo, nas Subseções 4.1 e 4.2, apresentamos as técnicas de regressão linear e regressão logística. Na Subseção 4.3 abordamos problemas comuns encontrados com algoritmos de *Machine Learning*, e na Subseção 4.4 apresentamos o procedimento de *Cross Validation*.

Para que não haja confusão, na Seção 5 definimos e ilustramos o que adotaremos por escoamento bifásico, padrão de escoamento e fração volumétrica.

Seguindo, na Seção 6 é feita a descrição da aplicação do que foi abordado nas seções anteriores.

Na Seção 7 são apresentados os experimentos numéricos realizados, e feitas algumas discussões a respeito dos resultados.

Por fim, na Seção 8 são apresentadas as considerações finais a respeito do trabalho.

2 Conceitos da teoria clássica de problemas inversos

Nesta Seção a menos que o contrário seja dito, adotaremos X e Y como espaços de Hilbert, e A como um operador linear e limitado. Além disso, para auxiliar na compreensão desta Seção o leitor pode verificar os Apêndices **A** e **B**, que possuem importantes definições e resultados utilizados a seguir.

Segundo **Kirsch (2011)**, podemos dizer que o problema direto é o valor de um operador A agindo sobre um elemento x conhecido, em um espaço modelo X , e o problema inverso como a solução da equação $Ax = b$. Isto é, o problema direto seria: dado x (e A), obter o valor de Ax ; já o problema inverso seria: dado b (e A), resolver $Ax = b$, para algum x .

Dentre os problemas inversos, existem os denominados problemas bem-postos, onde segundo a definição de Hadamard, um problema é bem-posto: se há uma solução para o problema (existência); se há no máximo uma solução para o problema (unicidade); se a solução depende continuamente dos dados (estabilidade).

Formalizando a definição de problema bem-posto, segue-se que:

Definição 1. (*Problema bem-posto*) Sejam X e Y espaços normados, $A : X \rightarrow Y$ uma função (linear ou não-linear). A equação $Ax = b$ é chamada de bem-posta, se valem as seguintes afirmações:

- *Existência:* para todo $b \in Y$, existe (ao menos um) $x \in X$, tal que $Ax = b$;
- *Unicidade:* para todo $b \in Y$ existe no máximo um x tal que $Ax = b$;
- *Estabilidade:* a solução x depende continuamente de b ; em outras palavras, para toda sequência (x_n) contida em X , com $A(x_n) \rightarrow A(x)$ ($n \rightarrow \infty$), segue que $x_n \rightarrow x$ ($n \rightarrow \infty$).

Equações as quais não valem uma das propriedades acima são chamadas de mal-postas.

Na tripla (X, Y, A) , é necessário especificar os espaços e suas normas. As questões de existência e unicidade dependem da natureza algébrica dos espaços e do operador. Já a estabilidade dos dados depende também da natureza topológica dos espaços, isto é, se $A^{-1} : Y \rightarrow X$ é uma função contínua.

2.0.1 Propriedades da Pseudo-inversa em espaços de Hilbert

A estabilidade dos dados, como visto anteriormente, depende da continuidade do operador inverso que está sendo trabalhado, no caso anterior o operador A , do problema

$Ax = b$. A seguir, vamos explorar um conceito de operador inverso mais generalizado, denominado pseudo-inversa.

Desejamos saber a solução da seguinte equação:

$$Ax = b$$

em que A é um operador linear e limitado, $A : X \rightarrow Y$, em que X e Y são espaços de Hilbert, e $b \in Y$. Se $b \notin R(A)$, isto é, se b não está no conjunto imagem do operador A , não existe solução para a equação. Neste caso desejamos encontrar os vetores que mais se aproximam da solução, isto é:

$$\|A\bar{x} - b\| = \inf_{x \in X} \|Ax - b\|$$

Se \bar{x} existe, este vetor é chamado *solução de mínimos quadrados* de $Ax = b$.

Para caracterizarmos a solução de mínimos quadrados, traremos o teorema a seguir, em que o operador adjunto de A , é denotado por $A^* \in \mathcal{L}(Y, X)$, e o operador de projeção sobre o fecho da imagem de A , por $P_{\overline{R(A)}}$.

Teorema 1. *Seja $b \in Y$ e $A \in \mathcal{L}(X, Y)$ com X e Y espaços de Hilbert. Dado $\bar{x} \in X$ as seguintes afirmações são equivalentes:*

- \bar{x} satisfaz $A\bar{x} = P_{\overline{R(A)}}b$;
- \bar{x} minimiza o resíduo: $\|A\bar{x} - b\| \leq \|Ax - b\|$ para todo $x \in X$;
- \bar{x} é uma solução das equações normais: $A^*A\bar{x} = A^*b$.

Prosseguindo na caracterização, definimos $S(b) := \{x \in X : A^*Ax = A^*b\}$, que é o conjunto das soluções das equações normais. Algumas das propriedades mais importantes desse conjunto, para o estudo que desejamos são:

- $S(b) \neq \emptyset$ se, e somente se, $b \in R(A) \oplus R(A)^\perp$;
- $S(b)$ é fechado e convexo;
- Se $b \in R(A) \oplus R(A)^\perp$, então existe um único x^\dagger , tal que:

$$x^\dagger \in S(b) : \|x^\dagger\| < \|\bar{x}\| \text{ para todo } \bar{x} \in S(b) \setminus \{x^\dagger\}$$

Tendo em vista o último teorema e propriedades, podemos definir o conceito de inversa generalizada, ou pseudo-inversa, do operador $A : X \rightarrow Y$, denotado por $A^\dagger : R(A) \oplus R(A)^\perp \subset Y \rightarrow X$ que, para cada b , associa o elemento $x^\dagger \in S(b)$. Em resumo, a imagem da pseudo-inversa é o vetor da solução de mínimos quadrados de $Ax = b$ que possui a menor norma.

Algumas das propriedades da pseudo-inversa, importantes para o presente trabalho:

- A^\dagger está definida em todo Y se, e somente se, $R(A)$ é fechada em Y ;
- A^\dagger é linear;
- A^\dagger é contínua se, e somente se, $R(A)$ é fechado em Y .

A partir daí definimos um problema mal-posto de outra forma.

Definição 2. (*Problema mal-posto*) Sejam X e Y espaços de Hilbert e $A : X \rightarrow Y$ um operador linear e contínuo. A tripla (A, X, Y) é chamada de problema mal-posto segundo Nashed, se $R(A)$ não é fechado em Y . Caso contrário, a tripla é definida como problema bem-posto segundo Nashed.

A última definição, implica que a tripla (A, X, Y) é um problema bem-posto segundo Nashed, se e somente se, a pseudo-inversa está definida e é contínua em todo Y .

Assim, no caso de um problema bem-posto, ocorre que $R(A)$ é fechado em Y , portanto A^\dagger é contínua (e portanto, limitada) e está definida em todo Y . Dessa forma, se $\|b^\delta - b\| \leq \delta$, como A^\dagger está definida em todo Y , existem x^δ e x^\dagger , tais que $A^\dagger b^\delta = x^\delta$, e $A^\dagger b = x^\dagger$. Dessa forma, temos:

$$\|x^\delta - x^\dagger\| = \|A^\dagger b^\delta - A^\dagger b\| = \|A^\dagger(b^\delta - b)\| \leq \|A^\dagger\| \|b^\delta - b\| \leq C\delta \text{ com } C > 0 \text{ real.}$$

Ou seja, se há continuidade nos dados, há continuidade na solução do problema inverso.

Caso contrário, isto é, se não tivermos a continuidade de A^\dagger , não conseguimos limitar a desigualdade anterior, pois não conseguimos garantir que exista $C > 0$ real, tal que $\|A^\dagger\| < C$.

2.0.2 Teoria da regularização

Nesta seção, para facilitar a compreensão do que é apresentado e nos preocuparmos com a estabilidade dos dados, deixaremos a pseudo-inversa de lado.

Vamos considerar X e Y espaços de Hilbert, $A : X \rightarrow Y$ um operador linear, injetivo e compacto¹, e que existem $x \in X, b \in Y$, de forma que $Ax = b$, ou seja, $b \in R(A)$. Podemos assumir a injetividade de A sem grandes problemas, pois caso o operador não seja injetivo, é possível tomar o espaço quociente $X/\mathcal{N}(A)$, em que $\mathcal{N}(A)$ é o núcleo de A , e tomar o operador induzido $\hat{A} : X/\mathcal{N}(A) \rightarrow Y$, que é injetivo (para mais detalhes, consultar Kirsch (2011)). Dessa forma, a injetividade de A nos garante que a solução é única.

Além disso, o fato do operador A ser linear e compacto, implica que se $\dim(X) = \infty$, então o problema $Ax = b$ será sempre mal-posto. (KIRSCH, 2011)

Retornando ao problema inverso, na prática nunca conhecemos exatamente o vetor $b \in Y$, da equação $Ax = b$. O que conhecemos normalmente é um vetor com ruído,

¹ Para mais detalhes sobre operadores compactos, consultar nos Apêndices da referência (KIRSCH, 2011).

digamos que seja um $b^\delta \in Y$ (não podemos garantir que $b^\delta \in R(A)$), com um nível de ruído $\delta > 0$, de forma que:

$$\|b - b^\delta\| \leq \delta$$

Como queremos resolver $Ax = b$, ou, mais geralmente, $A^{-1}b = x$, nos é necessário que tenhamos um vetor aproximado de x , digamos x^δ , mas também um requisito adicional a x^δ , é que dependa continuamente dos dados (b^δ). Isto é, dado que o operador A^{-1} é ilimitado, deseja-se encontrar uma aproximação para $A^{-1}b$, que é Rb^δ , em que $R : Y \rightarrow X$ é um operador linear e contínuo.

Com o intuito de encontrar uma aproximação para esse operador inverso não limitado, utilizaremos a teoria de regularização.

Definição 3. *Seja Λ um conjunto de índices. Uma estratégia de regularização para A é uma família de operadores lineares e limitados*

$$(R_\alpha)_{\alpha \in \Lambda} : Y \rightarrow X$$

tal que,

$$\lim_{\alpha \rightarrow 0} R_\alpha Ax = x \text{ para todo } x \in X.$$

Ou seja, os operadores $R_\alpha A$ convergem pontualmente para a identidade.

Considerando A um operador compacto e da definição anterior, alguns resultados importantes decorrem sobre a estratégia de regularização.

Teorema 2. *Seja $(R_\alpha)_{\alpha \in \Lambda}$ uma estratégia de regularização para o operador compacto $A : X \rightarrow Y$ em que a $\dim X = \infty$. Então, segue que:*

(1) *Os operadores R_α não são uniformemente limitados, isto é, existe uma sequência (α_j) convergente a zero, com $\|R_{\alpha_j}\| \rightarrow \infty$ quando $j \rightarrow \infty$.*

(2) *A sequência $(R_\alpha Ax)$ não converge uniformemente em conjuntos limitados de X , ou seja, não há convergência de $R_\alpha A$ para a identidade I na norma do operador.*

A estratégia de regularização se baseia a partir do problema inicial $Ax = b$, onde os dados são exatos, isto é, não são consideradas ruído nos dados. Também é o mesmo que dizer $R_\alpha b$ converge para x , considerando o valor exato de $b = Ax$.

Agora, considere $b \in R(A)$ como o valor exato da equação $b = Ax$, e $b^\delta \in Y$ o dado mensurado, com $\|b - b^\delta\| \leq \delta$. Definimos:

$$x^{\alpha, \delta} := R_\alpha b^\delta \tag{2.1}$$

como sendo uma aproximação da solução x . Assim, podemos mensurar o erro total, isto é, erro entre a solução exata e a estimada, através das seguintes desigualdades:

$$\|x^{\alpha, \delta} - x\| \leq \|R_\alpha b^\delta - R_\alpha b\| + \|R_\alpha b - x\|$$

$$\leq \|R_\alpha\| \|b^\delta - b\| + \|R_\alpha Ax - x\|$$

e portanto,

$$\|x^{\alpha,\delta} - x\| \leq \delta \|R_\alpha\| + \|R_\alpha Ax - x\|$$

A partir da desigualdade acima, podemos perceber que o erro total está limitado por dois termos; o primeiro é produto do ruído nos dados (δ) pela norma do operador de regularização R_α . Pelos resultados do último teorema, temos que $\|R_\alpha\| \rightarrow \infty$ quando $\alpha \rightarrow 0$. Já o segundo termo, se refere ao erro de aproximação o qual converge a zero quando α vai a zero, segundo a definição de R_α .

Pode-se ter uma noção de uma cota superior para o erro total conforme o seguinte gráfico para δ fixo:

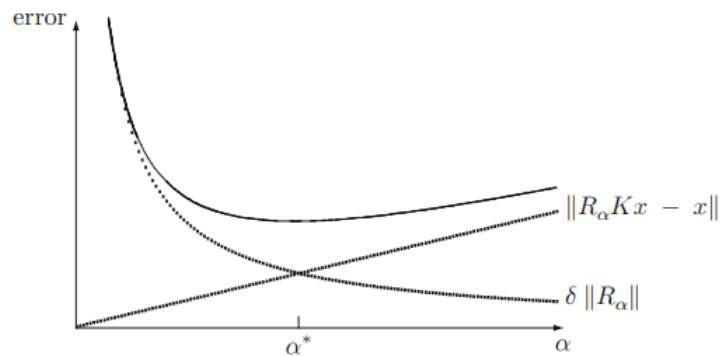


Figura 1 – Estimativa de cota superior para o erro total.

Fonte: KIRSCH (2011).

A partir disso, Kirsch afirma que em todo problema concreto ocorrerá o procedimento de estimar $\|R_\alpha\|$ e $\|R_\alpha Ax - x\|$, em termos de α e então, minimizar o erro total. Na Figura 1, o minimizador ocorre para o valor de α^* .

Logo, como a estimativa do erro total é necessária em todo problema. Introduzimos a seguinte definição:

Definição 4. Uma estratégia de regularização $\alpha = \alpha(\delta)$ é dita ser admissível se, $\alpha(\delta) \rightarrow 0$ e

$$\sup \{ \|R_{\alpha(\delta)} b^\delta - x\| : b^\delta \in Y, \|Ax - b^\delta\| \leq \delta \} \rightarrow 0, \delta \rightarrow 0$$

para todo $x \in X$.

2.0.2.1 Método Tikhonov

Nesta subseção as principais referências utilizadas foram [MARGOTTI](#) e [Kirsch \(2011\)](#).

No método de regularização de Tikhonov, novamente deseja-se resolver o problema $Ax = b$, sendo que conhecemos uma aproximação de b , isto é, $\|b - b^\delta\| \leq \delta$.

Vamos assumir que $A : X \rightarrow Y$ é um operador linear e contínuo, que age entre espaços de Hilbert, e também que o problema inverso $Ax = b$ possua uma única solução $x = x^*$.

Uma das maneiras de chegar a aproximação da solução (x^*), é minimizar o seguinte funcional:

$$f(x) = \frac{1}{2} \|Ax - b^\delta\|^2 \quad (2.2)$$

Porém, se o problema for mal-posto, podemos ter uma sequência $x_n \subset X$, em que ocorra: $Ax_n \rightarrow b$, mas $\|x_n\| \rightarrow \infty$.

Com isso, mesmo que tenhamos ruído (δ) pequeno, os vetores que minimizam f , podem estar muito distantes da solução exata. Em resposta a essa dificuldade, substituímos f , pelo funcional de *Tikhonov*, que procura resolver isso penalizando a norma do minimizador. O funcional de *Tikhonov* é definido a seguir:

Definição 5. Sejam X e Y espaços de Hilbert, $A : X \rightarrow Y$, um operador linear e limitado, $b^\delta \in Y$ e $\alpha > 0$. Define-se o Funcional de Tikhonov por:

$$T_\alpha(x) = \|Ax - b^\delta\|^2 + \alpha \|x\|^2 \text{ com } x \in X$$

Utilizamos o minimizador deste funcional como aproximação para solução do problema inverso. Ou seja:

$$x^\alpha \approx x^*, \text{ em que } x^\alpha = \operatorname{argmin}_{x \in X} T_\alpha(x) \quad (2.3)$$

O número real α é denominado de parâmetro de regularização. Sua escolha determina qual equilíbrio se terá entre precisão e estabilidade. Isso porque, se por um lado tomamos α muito pequeno, então penalizamos pouco a norma do minimizador x^α , o que implica

que a norma deste vetor será muito grande, dessa forma ficando normalmente distante da solução x^* . Por outro lado, se tomamos α grande, teremos maior estabilidade do funcional de Tikhonov, e x^α norma pequena, porém, estaremos dando pouca "importância" para minimizar o funcional do resíduo f , e assim x^α não representará bem a solução x^* .

O minimizador do funcional de Tikhonov x^α é único, conforme o teorema a seguir:

Teorema 3. *Sejam $A : X \rightarrow Y$ um operador linear e limitado entre espaços de Hilbert, $\alpha > 0$, $\delta > 0$ e $b^\delta \in Y$. Então, o funcional de Tikhonov $T_\alpha(x)$ tem um único mínimo $x^{\alpha,\delta} \in X$. Além disso, $x^{\alpha,\delta}$ é a única solução da equação:*

$$\alpha x^{\alpha,\delta} + A^* A x^{\alpha,\delta} = A^* b^\delta$$

Segundo Kirsch (2011), o minimizador x^α do teorema anterior, pode ser escrito da forma $x^{\alpha,\delta} = R_\alpha b^\delta$, sendo:

$$R_\alpha := (\alpha I + A^* A)^{-1} A^* : Y \rightarrow X \quad (2.4)$$

A equação (2.4) é a estratégia de regularização para o método de Tikhonov, que é admissível para valores apropriados de α pelo seguinte teorema:

Teorema 4. *Seja $A : X \rightarrow Y$ um operador linear e compacto, $\alpha > 0$ e $b^\delta \in Y$. Então o operador $\alpha I + A^* A$ é limitado e inversível. O operador $R_\alpha : Y \rightarrow X$ é denominado método de regularização de Tikhonov. $R_\alpha b^\delta$ é determinada pela única solução $x^{\alpha,\delta} \in X$ na equação:*

$$\alpha x^{\alpha,\delta} + A^* A x^{\alpha,\delta} = A^* b^\delta$$

Toda escolha $\alpha(\delta) \rightarrow 0$ ($\delta \rightarrow 0$) com $\frac{\delta^2}{\alpha(\delta)}$ ($\delta \rightarrow 0$), é admissível.

Do último resultado, vemos que para se obter um α admissível, deve-se escolher α que tenda a 0 conforme δ também tenda a zero, mas não tão rápido quanto δ^2 tende a 0.

2.0.2.2 Método tipo gradiente (Landweber)

Nesta parte abordaremos uma técnica de regularização que é o método do tipo gradiente. Nela, procuramos aproximar a solução do problema inverso:

$$Ax = b \quad (2.5)$$

com $\|b - b^\delta\| \leq \delta$, a partir de um ponto inicial, digamos x_0 , e seguir em uma direção apropriada a cada iteração, que veremos mais adiante.

Mas antes disso, precisamos apresentar alguns conceitos para melhor compreensão da técnica.

Por vezes, pode ser útil, saber a derivada de uma função em diversas direções, não somente nos vetores base de um espaço vetorial, como normalmente é apresentado na graduação. Por essa razão, segue o conceito de derivada direcional:

Definição 6. (*Derivada direcional*) Sejam X e Y espaços normados sobre o corpo dos reais, $U \subset X$ aberto e $F : U \rightarrow Y$. A derivada direcional de F no ponto $x \in U$ e na direção do vetor não nulo $v \in X$, é dada por:

$$DF(x)v = \lim_{t \rightarrow 0} \frac{F(x + tv) - F(x)}{t}$$

se este limite existir.

Pode-se demonstrar que a derivada direcional não é um operador linear, no que se refere as direções. Já a derivada de Gâteaux, possui essa propriedade:

Definição 7. (*Derivada de Gâteaux*) Sejam $F : D(F) \subset X \rightarrow Y$ um operador, X, Y espaços de Hilbert e $x \in \text{int}(D(F))$. Dizemos que F é G-diferenciável em x , se existe um operador linear e contínuo $A : X \rightarrow Y$ tal que:

$$DF(x)v = Av, \forall v \in X$$

Isto é, F é G-diferenciável em x , se existe um operador linear e contínuo que represente todas as derivadas direcionais de F no ponto x . Denotamos a derivada de Gâteaux em x por $F'(x)$. Se F é um funcional, isto é, se $Y = \mathbb{R}$, então escrevemos $\nabla F(x)$, e chamamos a derivada de Gâteaux de gradiente.

Vamos abordar alguns exemplos de derivadas de Gâteaux que serão úteis mais adiante.

Exemplo:

Seja X um espaço de Hilbert real. Considere a função $F : X \rightarrow \mathbb{R}$, com $F(x) = \frac{1}{2}\|x\|^2$.

Fixe $x \in X$. Vamos verificar se F é G-diferenciável em x .

Seja $v \in X$. Sabemos que a definição de derivada direcional de F no ponto x na direção de v é dada por:

$$\begin{aligned} DF(x)v &= \lim_{t \rightarrow 0} \frac{F(x + tv) - F(x)}{t} = \lim_{t \rightarrow 0} \frac{\frac{1}{2}\|x + tv\|^2 - \frac{1}{2}\|x\|^2}{t} = \\ &= \lim_{t \rightarrow 0} \frac{\frac{1}{2}(\|x\|^2 + 2\langle x, tv \rangle + \|tv\|^2) - \frac{1}{2}\|x\|^2}{t} = \lim_{t \rightarrow 0} \frac{\frac{1}{2}(2t\langle x, v \rangle + t^2\|v\|^2)}{t} = \\ &= \lim_{t \rightarrow 0} \frac{t(2\langle x, v \rangle + t\|v\|^2)}{t} = \lim_{t \rightarrow 0} (\langle x, v \rangle + \frac{t}{2}\|v\|^2) = \langle x, v \rangle \end{aligned}$$

Portanto, temos que a derivada direcional de F no ponto x na direção de v é $DF(x)v = \langle x, v \rangle$.

Com isso podemos definir o seguinte operador $A_x : X \rightarrow \mathbb{R}$, $v \mapsto \langle x, v \rangle$. Conseguimos ver que $DF(x)v = A_x v = \langle x, v \rangle$, e assim pode-se identificar o A_x , com o vetor x e também com a derivada de Gâteaux F no ponto x , $A_x = x = F'(x)$, isto é, $F'(x)v = \langle x, v \rangle$. De fato, concluímos que F é G-diferenciável em x , dado que A_x é linear e contínua, pelas

propriedades do produto interno e pela desigualdade de Cauchy-Schwartz.

Outro exemplo:

Seja $A : X \rightarrow Y$ um operador linear e contínuo, X, Y espaços de Hilbert reais, e $b \in Y$. Considere o funcional $F : X \rightarrow \mathbb{R}$, com $F(x) = \frac{1}{2}\|Ax - b\|^2$. De forma análoga a anterior, temos:

$$\begin{aligned} DF(x)v &= \lim_{t \rightarrow 0} \frac{F(x + tv) - F(x)}{t} = \lim_{t \rightarrow 0} \frac{\frac{1}{2}(\|A(x + tv) - b\|^2 - \|Ax - b\|^2)}{t} = \\ &= \lim_{t \rightarrow 0} \frac{\frac{1}{2}(\|Ax - b + Atv\|^2 - \|Ax - b\|^2)}{t} = \lim_{t \rightarrow 0} \frac{\frac{1}{2}(\|Ax - b\|^2 + 2\langle Ax - b, Atv \rangle + \|Atv\|^2 - \|Ax - b\|^2)}{t} = \\ &= \lim_{t \rightarrow 0} \frac{\frac{1}{2}(2\langle Ax - b, Atv \rangle + \|Atv\|^2)}{t} = \lim_{t \rightarrow 0} \frac{\frac{1}{2}(2t\langle Ax - b, Av \rangle + t^2\|Av\|^2)}{t} = \\ &= \lim_{t \rightarrow 0} (\langle Ax - b, Av \rangle + \frac{1}{2}t\|Av\|^2) = \langle Ax - b, Av \rangle \end{aligned}$$

Por fim, da propriedade do operador adjunto em espaços de Hilbert, temos que:

$$DF(x)v = \langle Ax - b, Av \rangle = \langle A^*(Ax - b), v \rangle$$

e portanto, podemos identificar a derivada de Gâteaux de F no ponto x com:

$$F'(x) = A^*(Ax - b). \quad (2.6)$$

Definição 8. (*Direção de descida*) Sejam X um espaço de Hilbert, $f : X \rightarrow \mathbb{R}$ um funcional e $x \in X$. Dizemos que $v \in X$ é direção de descida do funcional f a partir do ponto x se existe $\bar{t} > 0$ tal que, $f(x + tv) < f(x)$ para todo $t \in \mathbb{R}$ tal que $0 < t \leq \bar{t}$.

Como mencionado anteriormente, partindo de um ponto arbitrário, a cada iteração no método do gradiente, seguimos em uma direção adequada. Esta direção tem o intuito de minimizar o seguinte funcional:

$$f(x) = \frac{1}{2}\|Ax - b^\delta\|^2 \quad (2.7)$$

que se torna equivalente a encontrar uma aproximação para $A^\dagger b^\delta$.

Com a definição anterior, se escolhermos uma direção de descida para f , e tomarmos um passo suficientemente pequeno nesta direção, estaremos reduzindo o valor do funcional. Mas como encontramos a direção de descida? O seguinte teorema nos mostra uma alternativa:

Teorema 5. *Sejam X um espaço de Hilbert, $f : X \rightarrow \mathbb{R}$ um funcional, $x \in X$, f G-diferenciável em x e $\nabla f(x) \neq 0$. Então $v = -\nabla f(x)$ é direção de descida de f a partir de x .*

Demonstração. Sejam $x, w \in X$, temos que a derivada direcional de f na direção de w a partir do ponto x é $\nabla f(x)w = Df(x)w = \lim_{t \rightarrow 0} \frac{f(x+tw) - f(x)}{t}$.

Sabemos que por f ser G-diferenciável em x , podemos escrever $\nabla f(x)v = \langle \nabla f(x), v \rangle$. Logo, se tomarmos $v = -\nabla f(x)$, temos que $\nabla f(x)v = -\|\nabla f(x)\|^2 < 0$.

Mas também teremos que, se tomarmos o limite pela direita que: $\nabla f(x)v = \lim_{t \rightarrow 0^+} \frac{f(x+tv) - f(x)}{t} < 0$. O que implica que existe $\bar{t} > 0$ tal que para todo t , com $0 < t \leq \bar{t}$, vale $f(x+tv) < f(x)$. \square

Ainda se pode mostrar que não só a direção oposta ao gradiente é uma direção de descida, como é a direção na qual o funcional decresce mais rápido a partir do ponto escolhido.

Tendo em vista o exposto, podemos explicar o método do gradiente. Neste tipo de método, na n -ésima iteração (cujo o ponto respectivo consideramos x_n), o gradiente do resíduo que denotamos por $\nabla f(x_n)$, é calculado no ponto x_n . Assim, como a direção oposta a do gradiente é uma direção de descida, se tomarmos um valor relativamente pequeno $\lambda_n > 0$, teremos que $x_n - \lambda_n \nabla f(x_n)$, tem resíduo menor do que a do x_n .

Assim, a iteração de um método do tipo gradiente pode ser definida da seguinte forma:

- escolha um ponto inicial $x_0 \in X$ (um chute), e defina $n = 0$;
- escolha o tamanho do passo $\lambda_n > 0$;
- defina $x_{n+1} = x_n - \lambda_n \nabla f(x_n)$;
- substitua $n + 1$ por n e repita a partir do passo 2.

O gradiente do funcional em (7), num dado ponto x qualquer, é representado por (2.6). Assim, cada iteração do método do gradiente pode ser escrita da seguinte forma:

$$x_{n+1} = x_n - \lambda_n A^*(Ax_n - b^\delta). \quad (2.8)$$

Vale lembrar que o valor de λ_n deve ser pequeno o suficiente para que x_{n+1} possua um resíduo menor que x_n , conforme a definição de direção de descida. Cada método possui uma maneira diferente de estabelecer o tamanho do passo λ .

Tendo introduzido um conceito geral de métodos do tipo gradiente, no presente trabalho utilizaremos o *Método de Landweber*, que consiste em escolher um tamanho de passo fixo λ , com determinadas restrições.

3 Tomografia por Impedância Elétrica

Nessa seção descrevemos um problema inverso de interesse prático, a assim chamada Tomografia por Impedância Elétrica (EIT na abreviatura em inglês). Os detalhes desse modelo são explicados na referência [Borcea \(2002\)](#) para a Subseção 4.1, e [Somersalo, Cheney e Isaacson \(1992\)](#) para a Subseção 4.2.

Neste problema deseja-se saber a condutividade elétrica no interior de um corpo de prova através da aplicação de correntes elétricas na fronteira do mesmo, e medindo-se a diferença de potencial também na fronteira. Uma das aplicações da Tomografia por Impedância Elétrica é utilizar essa condutividade para reconstruir a imagem do que está no interior do corpo de prova. Existem várias formas de modelar este problema. Vamos apresentar duas delas: o modelo contínuo e o modelo completo de eletrodos.

3.0.1 O Modelo Contínuo

Vamos identificar o corpo de prova com o conjunto aberto $\Omega \subset \mathbb{R}^2$, as correntes elétricas com $g : \partial\Omega \rightarrow \mathbb{R}$, que são aplicadas na fronteira de Ω , a qual é denotada por $\partial\Omega$.

Seja $\gamma : \Omega \rightarrow \mathbb{R}$ a condutividade elétrica de Ω . Assume-se que Ω não possui fontes ou drenos de fluxo, o que implica que o fluxo elétrico $-\gamma\nabla u$ tem divergente nulo, onde $u : \Omega \rightarrow \mathbb{R}$ representa a distribuição de potencial elétrico em Ω . Essa condição pode ser escrita como:

$$\operatorname{div}(\gamma\nabla u) = 0 \text{ em } \Omega \quad (3.1)$$

Assume-se que o fluxo elétrico na fronteira, é exatamente a corrente elétrica medida, o que implica em:

$$-\gamma\langle\nabla u, \nu\rangle = g \text{ em } \partial\Omega. \quad (3.2)$$

em que ν é o vetor normal unitário exterior ao conjunto Ω .

Com o objetivo de provar existência e a unicidade de uma solução u para a EDP determinada pelas equações (3.6) e (3.2), podemos substituir as equações pela formulação fraca, como apresentado em [Margotti \(2015\)](#), isto é:

Seja $g \in H_{\diamond}^{-1/2}(\partial\Omega)^1$ e $\gamma \in L_{+}^{\infty}(\Omega)$, deseja-se encontrar uma função $u \in H_{\diamond}^1(\Omega)$, de forma que:

$$\int_{\Omega} \gamma\nabla u\nabla\varphi = \int_{\partial\Omega} g\varphi \text{ para toda } \varphi \in H_{\diamond}^1(\Omega). \quad (3.3)$$

A função φ é chamada de função teste. Já o símbolo \diamond significa que a integral da função sobre a fronteira de Ω é nula:

¹ H denota um espaço de Sobolev, que é um espaço de Hilbert, com algumas características específicas. Para mais informações pode-se consultar ([KIRSCH, 2011](#)).

$$H_{\diamond}^1(\Omega) := \left\{ u \in H^1(\Omega) : \int_{\partial\Omega} u = 0 \right\}.$$

Se a função é definida em Ω , esta integral é entendida no sentido do teorema do traço: para $u \in H^1(\Omega)$, seu traço² $f = u|_{\partial\Omega}$ pertence a $H^{1/2}(\partial\Omega)$, e define-se:

$$H_{\diamond}^{1/2}(\partial\Omega) = \left\{ v \in H^{1/2}(\partial\Omega) : \int_{\partial\Omega} v = 0 \right\}$$

O conjunto $H_{\diamond}^{-1/2}(\partial\Omega)$ é definido como o espaço dual de $H_{\diamond}^{1/2}(\partial\Omega)$ e:

$$L_{+}^{\infty}(\Omega) := \{v \in L^{\infty}(\Omega) : v \geq C \text{ quase sempre em } \Omega\}$$

em que $C > 0$ é uma constante.

A condição $\int_{\partial\Omega} g = 0$ ($g \in H_{\diamond}^{-1/2}(\partial\Omega)$) é interpretada como a lei de conservação de cargas. Esta, em conjunto com o lema de Lax-Milgram, é utilizada para provar a existência de uma solução $u \in H^1(\Omega)$ para a formulação fraca (11) (única, a menos de uma constante). Já a condição que garante a unicidade da solução é dada por $\int_{\partial\Omega} f = 0$ ($u \in H_{\diamond}^1(\Omega)$), que é interpretada como o referencial para a diferença de potencial. Além disso, o limite inferior $\gamma \geq C$ em Ω ($\gamma \in L_{+}^{\infty}(\Omega)$) garante que o fluxo elétrico pode percorrer o conjunto Ω .

Garantida a existência e unicidade da solução de (3.3), o operador Neumann-Dirichlet (NtD), denotado por Λ_{γ} , que associa a corrente elétrica g com a respectiva voltagem produzida f está bem definido:

$$\Lambda_{\gamma} : H_{\diamond}^{-1/2}(\partial\Omega) \rightarrow H_{\diamond}^{1/2}(\partial\Omega), g \mapsto f \quad (3.4)$$

Este operador é linear e contínuo. Além disso, possui um operador inverso linear e limitado, chamado de operador Dirichlet-Neumann.

O operador direto associado ao problema de TIE no modelo contínuo, é definido pela função não-linear:

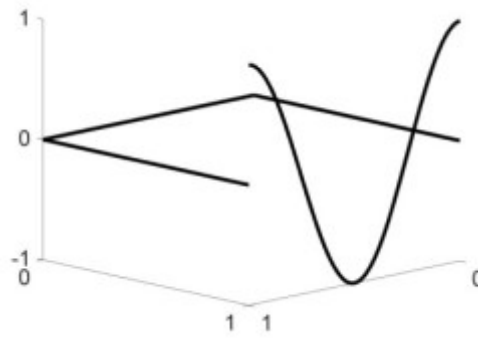
$$F : D(F) \subset L^{\infty}(\Omega) \rightarrow \mathcal{L}(H_{\diamond}^{-1/2}(\partial\Omega), H_{\diamond}^{1/2}(\partial\Omega)), \gamma \mapsto \Lambda_{\gamma} \quad (3.5)$$

Sendo $D(F) := L_{+}^{\infty}(\Omega)$. Daí, o problema inverso da TIE no modelo contínuo, consiste em reconstruir a condutividade elétrica γ , a partir de informações do operador Neumann-Dirichlet, que é um problema não linear e mal-posto.

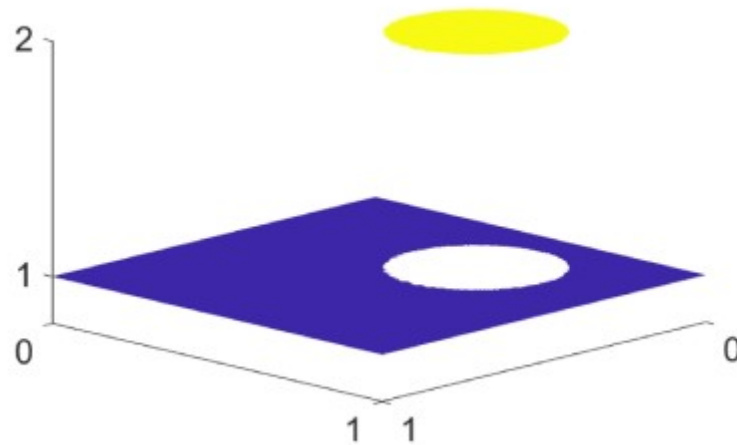
Agora, vamos apresentar um exemplo do que ocorre num experimento do modelo contínuo. Neste exemplo, Ω é o quadrado $(0, 1) \times (0, 1) \subset \mathbb{R}^2$, e g é uma corrente aplicada em $\partial\Omega$, como vemos na Figura 2.

Escolhemos um exemplo de condutividade γ em Ω , de forma a representar um círculo com condutividade 2, e o restante com condutividade 1, conforme a Figura 3.

² Para mais informações sobre o teorema do traço consultar Brezis (2010).

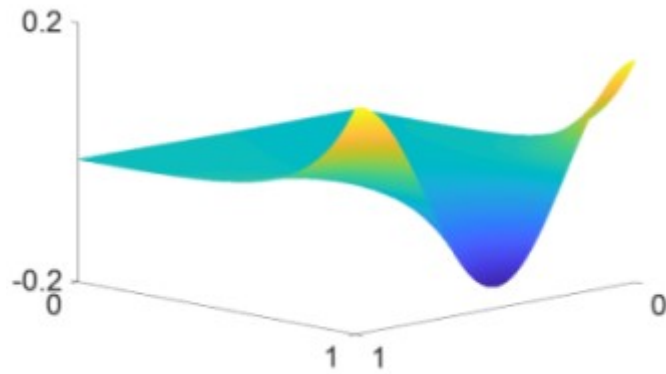
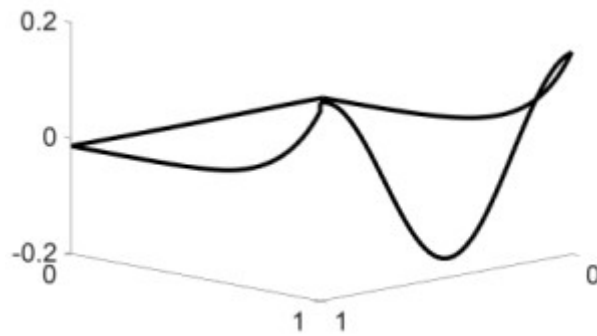
Figura 2 – Uma das correntes g aplicadas em $\partial\Omega$.

Fonte: Fábio Júnior Margotti.

Figura 3 – Exemplo de uma condutividade γ aplicada em Ω .

Após a aplicação da corrente g , é gerada uma função potencial u em Ω , a qual representa a solução de (11). Esta função (reconstruída computacionalmente através do Método de Elementos Finitos) está plotada na Figura 4.

Por fim, temos as voltagens mensuradas em $\partial\Omega$, que nada mais são que u restrita a fronteira de Ω . Esta é a função f que é apresentada na Figura 5.

Figura 4 – Função potencial $u : \Omega \rightarrow \mathbb{R}$.Figura 5 – Função u restrita a fronteira, ou seja, f .

3.0.2 O Modelo Completo de Eletrodos

Este modelo parte do mesmo princípio do apresentado na seção anterior. Porém agora, ao invés de aplicarmos as correntes elétricas em toda a fronteira de Ω , aplicaremos apenas em eletrodos posicionados em $\partial\Omega$. Adotamos o modelo de eletrodos, pois, apesar do modelo contínuo ter algumas facilidades matemáticas, conforme afirma [Somersalo, Cheney e Isaacson \(1992\)](#), o modelo completo de eletrodos traz melhor aproximação aos experimentos reais. Esse novo modelo acarreta algumas mudanças nas condições de contorno da EDP apresentadas na seção anterior.

Antes das novas condições, vale ressaltar que ainda consideramos que o corpo de prova Ω , não possui fontes ou drenos de fluxo elétrico, e portanto, assim como na subseção anterior, temos que o divergente do fluxo elétrico é nulo:

$$\operatorname{div}(\gamma \nabla u) = 0 \text{ em } \Omega, \quad (3.6)$$

com $u : \Omega \rightarrow \mathbb{R}$ o potencial elétrico e $\gamma : \Omega \rightarrow \mathbb{R}$, a condutividade elétrica.

Agora, seja L o número de eletrodos acoplados em $\partial\Omega$. Identificamos estes eletrodos como subconjuntos da fronteira, denotados por e_i , com $1 \leq i \leq L$, e assim $e_i \subset \partial\Omega$. Consideramos que não há sobreposição de eletrodos, portanto, $\bar{e}_i \cap \bar{e}_j = \emptyset$, se $i \neq j$.

Supomos que só conhecemos uma média da corrente elétrica em cada um dos eletrodos, ou seja,

$$\int_{e_i} \gamma \langle \nabla u, \nu \rangle dS = I_i, \text{ com } i = 1, 2, \dots, L, \quad (3.7)$$

em que $I = (I_1, I_2, \dots, I_L) \in \mathbb{R}^L$ é um vetor conhecido, e ν é o vetor normal unitário, saindo de Ω . Fora dos eletrodos a corrente elétrica é nula, isto é:

$$\gamma \langle \nabla u, \nu \rangle = 0, \text{ fora de } \bigcup_{i=1}^L e_i. \quad (3.8)$$

Consideramos que o metal dos eletrodos é um condutor perfeito, o que implica que o potencial elétrico em cada eletrodo é constante. Mas também levamos em conta, um fator denominado de impedância de contato, que representa uma fina camada de alta resistência elétrica entre o eletrodo e Ω . Denotamos por $z_i > 0$ a impedância de contato relativa ao eletrodo e_i . A queda de voltagem devida a essa pequena camada é dada pelo produto da impedância de contato pela corrente elétrica. Assim, temos que:

$$u + z_i \gamma \langle \nabla u, \nu \rangle = U_i, \text{ com } i = 1, 2, \dots, L. \quad (3.9)$$

Definimos $U = (U_1, U_2, \dots, U_L) \in \mathbb{R}^L$.

Ainda é necessário considerar a conservação de cargas, dada por:

$$\sum_{i=1}^L I_i = 0 \quad (3.10)$$

que é necessária para garantir a existência de solução. Já a condição para as voltagens é dada por:

$$\sum_{i=1}^L U_i = 0. \quad (3.11)$$

A condição em (3.11) é requisito para garantir a unicidade da solução.

Garantida existência e unicidade da solução, o seguinte operador Neumann para Dirichlet está bem definido:

$$\Lambda_\gamma : \mathbb{R}_\diamond^L \rightarrow \mathbb{R}_\diamond^L, \quad I \mapsto U \quad (3.12)$$

em que $\mathbb{R}_\diamond^L = \left\{ V = [v^{(1)}, v^{(2)}, \dots, v^{(L)}] \in \mathbb{R}^L : \sum_{i=1}^L v^{(i)} = 0 \right\}$

O operador direto associado ao modelo completo de eletrodos é definido por:

$$F : L_+^\infty(\Omega) \subset L^\infty(\Omega) \rightarrow \mathcal{L}(\mathbb{R}_\diamond^L, \mathbb{R}_\diamond^L), \quad \gamma \mapsto \Lambda_\gamma \quad (3.13)$$

Agora, resumidamente, no problema direto conhecemos γ , e para cada corrente I aplicada, queremos determinar U . Já no problema inverso conhecemos pares (I, U) e queremos determinar γ .

Na prática fixamos n correntes em $\mathbb{R}_{\diamond}^L \{I^1, I^2, \dots, I^n\}$, e redefinimos o operador F da seguinte forma:

$$F : L_+^{\infty}(\Omega) \subset L^{\infty}(\Omega) \rightarrow (\mathbb{R}_{\diamond}^L)^n, \gamma \mapsto (U^1, U^2, \dots, U^n) \quad (3.14)$$

em que $U^j = \Lambda_{\gamma}(I^j)$, com $j = 1, 2, \dots, n$.

4 Conceitos e técnicas básicas de *Machine Learning*

Nesta seção a principal referência utilizada será a do curso de *Machine Learning* do autor NG. Entretanto, outras referências recomendadas para este assunto são: Deisenroth, Faisal e Ong (2020) e Géron (2017).

Segundo o autor NG uma definição antiga de *Machine Learning*, dada por Arthur Samuel é: “o campo de conhecimento que fornece ao computador a habilidade de aprender sem ser explicitamente programado.”.

Uma definição mais atual, é dada por Mitchell (1997) : “Se diz que um programa de computador aprende da experiência E, em relação a alguma classe de tarefas T e medida de performance P, quando sua performance na tarefa T, mensurada por P, melhora com a experiência E”. Apesar desse aprendizado não ser explicitamente programado, existem os algoritmos de aprendizagem para as técnicas de *Machine Learning*, sendo basicamente divididos em: Algoritmo de Aprendizagem Supervisionada (*Supervised Learning Algorithm*) e Algoritmo de Aprendizagem Não-Supervisionada (*Unsupervised Learning*).

Os Algoritmos de Aprendizagem Não-Supervisionada têm com experiência uma base de dados contendo diversas características, e procuram aprender as características principais dessa estrutura de dados. Um exemplo disso é o chamado *clustering*, onde o algoritmo, a partir da base dados, agrupa-os conforme suas similaridades.

Os Algoritmos de Aprendizagem Supervisionada têm acesso a uma base de dados, contendo diversas características, mas cada exemplo está associado a alguma(s) característica(s) específica(s). Por exemplo, um conjunto de e-mails contendo Spams e e-mails normais, é utilizado para treinamento de um algoritmo, que em seguida classifica os próximos e-mails recebidos como Spam ou não.

Apesar da diversa gama de algoritmos, no presente trabalho estamos interessados em dois Algoritmos de Aprendizagem Supervisionada: a Regressão Linear e a Regressão Logística.

Para o treinamento dos algoritmos é necessário um conjunto de pares de dados, que chamamos de conjunto de treinamento. Para estabelecer notação para as seguintes subseções, denotaremos o conjunto de treinamento com m exemplos, como o conjunto de m pares $(x^{(i)}, y^{(i)})$, em que $x^{(i)}$ são as variáveis de entrada, e $y^{(i)}$ são as variáveis de saída, sendo que o índice i ($1 \leq i \leq m$), indica qual é o número do exemplo. Utilizando nosso contexto sobre classificação de e-mail, tomemos um conjunto de 10 exemplos, $x^{(1)}$ seria um e-mail, e $y^{(1)}$ seria sua respectiva classificação de Spam ou não, e assim se seguiria para os demais 9 e-mails.

Agora, descrevendo de maneira um pouco mais formal um Algoritmo de Aprendizagem

gem Supervisionada, o objetivo é: dado um conjunto de treinamento, o algoritmo deve aprender a função $h : X \rightarrow Y$, para que $h(x)$ seja uma boa previsão para o valor correspondente y . Segundo Ng, por razões históricas esta função h é chamada de *hipótese*. Podemos esquematizar este procedimento da seguinte maneira:

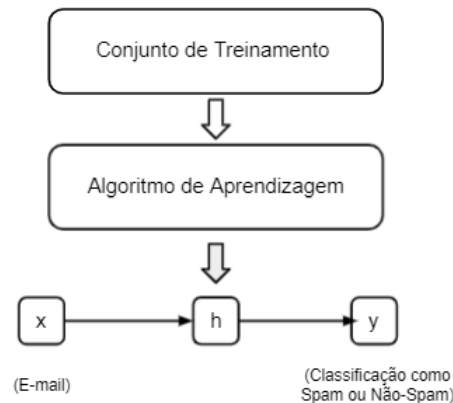


Figura 6 – Exemplo de funcionamento um Algoritmo de Aprendizagem Supervisionado.

Quando a variável que queremos prever através do algoritmo é discreta, como no caso da classificação de e-mails, chamamos de problema de classificação e utilizamos a técnica de regressão logística. Já se a variável for contínua, chamamos de problema de regressão e utilizamos a técnica de regressão linear.

Por fim, para avaliar se um algoritmo aprendeu, isto é, se a função hipótese é efetiva em prever os resultados, utilizamos um conjunto de dados não utilizados no treinamento, chamado conjunto teste. Neste conjunto, da mesma forma que no de treinamento, conhecemos pares de dados (pares (x, y)), porém, nas informações de entrada aplicamos a função hipótese, e comparamos a previsão com o resultado já conhecido. Assim podemos ver se a previsão do algoritmo está dentro dos parâmetros desejados ou não, para então podermos avaliar se o algoritmo aprendeu ou não.

4.0.0.1 Regressão Linear

A regressão linear é o processo de estimar os parâmetros de uma reta que se aproxime ao máximo de um conjunto de pontos no plano. Por exemplo, dado um conjunto de m pontos $(x^{(i)}, y^{(i)}) \in \mathbb{R} \times \mathbb{R}$, $i = 1, 2, \dots, m$, quais são os parâmetros a e b , da reta r , em que $r(x) = ax + b$, tais que a função $F(a, b)$ atinja o mínimo.

$$F(a, b) = \sum_{i=1}^m |r(x^{(i)}) - y^{(i)}|$$

Já no caso de várias dimensões, digamos n -dimensional, com $n \in \mathbb{N}$, dado um conjunto de m pontos $(x^{(i)}, y^{(i)}) \in \mathbb{R}^{n-1} \times \mathbb{R}$, $i = 1, 2, \dots, m$, queremos encontrar o vetor $[\theta_0, \theta]$,

de forma que o hiperplano n -dimensional h_θ , definido por $h_\theta(x) = \theta_0 + \theta \cdot x$, satisfaça $h_\theta(x^{(i)}) \approx y^{(i)}$, $i = 1, 2, \dots, m$. De forma semelhante ao problema anterior, deseja-se encontrar agora θ_0 e θ de forma a minimizar:

$$J(\theta) = \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2 \quad (4.1)$$

Adaptando o problema para a linguagem de *Machine Learning*: dado um conjunto de treinamento com m exemplos, com os pares $(x^{(i)}, y^{(i)}) \in \mathbb{R}^{n-1} \times \mathbb{R}$, $1 \leq i \leq m$, desejamos treinar o algoritmo de regressão linear para encontrar a função hipótese $h_\theta(x) = \theta_0 + \theta x$ ($\theta_0 \in \mathbb{R}$ e $\theta \in \mathbb{R}^{n-1}$), de forma que, dado x , obteremos uma predição feita por h_θ , isto é, $h_\theta(x)$. O treinamento do algoritmo é feito através da minimização da função que denominamos *função custo*, que é similar a equação do exemplo anterior:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2 \quad (4.2)$$

4.0.0.2 Regressão Logística Simples

Na regressão logística simples estamos interessados em que a máquina aprenda a classificar uma informação entre duas opções. Como por exemplo, a classificação de um e-mail como Spam ou como Não-Spam.

O treinamento do problema de classificação é similar ao de regressão linear, porém agora queremos que o algoritmo nos forneça valores discretos, ao invés de valores contínuos. No caso dos e-mails, podemos associar o valor 0 para Não-Spam e 1 para Spam. Para que isso ocorra, substituímos os valores de saída ($y^{(i)}$) por 0 para exemplos de uma classificação, e por 1 para os exemplos que possuem outra classificação, e assim esperamos que o algoritmo nos forneça predições com valores 0 ou 1.

Aparentemente, devido apenas a mudança dos valores saída, poderíamos utilizar a mesma função hipótese da regressão linear, para que fosse realizado o treinamento, mas **NG** afirma que utilizá-la normalmente trás resultados não satisfatórios. Para classificação, **NG** sugere utilizarmos outra função hipótese, que retorna valores entre 0 e 1, que seriam as duas classificações desejadas. Esta função hipótese é dada por:

$$h_\theta(x) = g(\theta^T x) \quad (4.3)$$

Em que:

$$g(z) = \frac{1}{1 + e^{-z}} \quad (4.4)$$

A função $g : \mathbb{R} \rightarrow \mathbb{R}$ é denominada *função sigmoide*. O gráfico da Figura 7 ilustra o comportamento desta função.

Aqui consideramos um conjunto de treinamento com m exemplos, com os pares $(x^{(i)}, y^{(i)}) \in \mathbb{R}^n \times \mathbb{R}$, $1 \leq i \leq m$, sendo $x^{(i)} = (1, x_1^{(i)}, x_2^{(i)}, \dots, x_{n-1}^{(i)})$, $\theta \in \mathbb{R}^n$ e $z \in \mathbb{R}$.



Figura 7 – Gráfico da função sigmóide.

Dado que os valores de (4.4) estão entre 0 e 1, podemos interpretar os valores dados pela função hipótese, como a probabilidade da predição ser 1. Isto é, se obtivermos $h_{\theta}(x) = 0,7$, isso representa a probabilidade de 70% da predição de x ser classificada como 1. Devido a esta probabilidade, no momento do treinamento do algoritmo, precisamos definir um critério para a predição do algoritmo, conforme o valor da função sigmoide encontrada. Por exemplo, podemos adotar o critério de que se $h_{\theta}(x) \geq 0.5$ então $y = 1$ e $h_{\theta}(x) < 0.5$ implica $y = 0$.

Novamente, para estimar o parâmetro θ , poderíamos segundo NG, utilizar a mesma função custo da regressão logística, mas dessa forma teríamos uma função não convexa, ou seja, uma função com possivelmente vários mínimos locais.

Para contornar isso utilizamos a seguinte função custo:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))] \quad (4.5)$$

No caso em que seja necessária a classificação entre mais de duas opções, optamos por uma estratégia um pouco diferente, mas utilizando as mesmas ferramentas anteriores. Para três ou mais classificações distintas, chamamos a técnica de regressão linear múltipla ou *um contra todos*. Nessa técnica, isolamos uma opção por vez e alteramos seus valores de saída para 1. Às demais opções adotamos os valores de saída como sendo 0, aplicamos a técnica de regressão linear simples e salvamos a função hipótese treinada. Repetimos o procedimento com todas as opções de classificação, e salvamos todas as funções hipóteses. Por fim, para o teste, aplicamos cada uma das funções hipótese em todos os dados de teste, e salvamos as respostas, que serão valores entre 0 e 1, devido a característica da função sigmoide. Adotaremos como a predição correta, a opção cuja função hipótese teve o maior valor.

4.0.0.3 High Bias e Overfitting

Bias é o que ocorre quando o algoritmo de *Machine Learning* não consegue estabelecer uma real relação entre os dados fornecidos, isto é, o algoritmo está pressupondo relações falsas. High Bias ocorre quando o erro na predição é muito impactante no algoritmo de *Machine Learning*, provocando o que chamamos de *underfitting*. *Underfitting* ocorre

quando quando as previsões do algoritmo não correspondem à realidade do conjunto de dados fornecidos, tanto do treinamento quanto do teste.

Segundo NG, para evitarmos o underfitting, podemos: adicionar variáveis (dimensão dos vetores $x^{(i)}$) ou aumentar a complexidade da função hipótese.

Já no overfitting, o que ocorre é que a função hipótese encaixa muito bem nos dados para o conjunto de dados de treinamento, dessa forma apresenta uma taxa de erro muito baixa neste conjunto. Porém, as previsões do algoritmo com dados novos (dados do conjunto teste), não “encaixam”, isto é, a taxa de erro é muito alta.

Segundo NG, para evitarmos o overfitting, podemos: usar mais exemplos no conjunto de treinamento ou diminuir o número de variáveis.

Para exemplificar o que acabou de ser apresentado, suponha que estamos tentando treinar um algoritmo para prever o preço de uma casa a partir de seu tamanho. Toma-se então um conjunto de casas, com seus tamanhos e respectivos preços. Representamos cada casa como um ponto no plano cartesiano, com o eixo das abscissas representando o tamanho das casas, e o eixo das ordenadas o preço.

Na Figura 8 são apresentados três gráficos, onde a curva em azul é a função hipótese obtida após o treinamento do algoritmo, e os ‘xis’ em vermelho são as marcações no plano cartesiano dos tamanhos das casas e seus respectivos preços (representando o conjunto de treinamento). No primeiro gráfico, vemos um exemplo de *High Bias*, pois a função hipótese não se encaixa bem ao dados de treinamento, nem consegue ter uma boa predição. No segundo gráfico vemos o que seria uma função hipótese ideal, pois consegue prever bem os dados. Já no terceiro gráfico temos um exemplo de *overfitting*, pois a função hipótese se encaixa muito bem nos dados de treinamento, mas é muito complexa e provavelmente apresentará predições ruins quando aplicada a dados ainda não vistos.

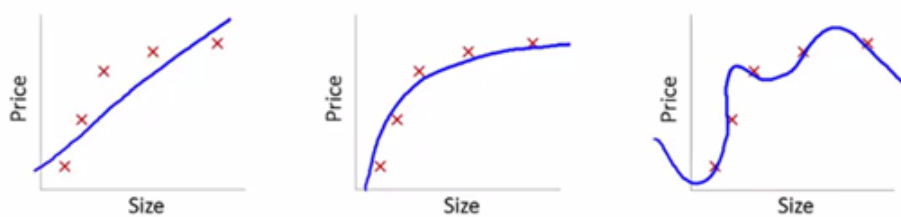


Figura 8 – Exemplos de gráficos funções hipótese com *Underfitting*, com boa predição de dados e com *Overfitting*, respectivamente.

Fonte: Curso de *Machine Learning* do autor NG.

4.0.1 Cross Validation

Cross Validation é um método de reamostragem dos dados para avaliar a capacidade de um algoritmo de generalizar, ou seja, através de uma reconfiguração dos dados utilizados

para o treinamento do algoritmo, quão bem o algoritmo consegue realizar previsões a partir de novas informações (BERRAR, 2018).

Existem diversas maneiras de realizar o *Cross Validation*, como pode ser visto em (BERRAR, 2018).

Aqui exploraremos a forma do autor NG, de realizar o *Cross Validation*. Este método tem o intuito de selecionar a função hipótese h que tem melhor capacidade de generalizar a previsão de dados. Suponhamos que temos um conjunto de dados com 1000 exemplos, isto é, 1000 pares de dados $(x, y) \subset \mathbb{R}^2$. Para realizar o *Cross Validation*, iremos subdividir o conjunto de 1000 exemplos em: 600 exemplos para o conjunto de treinamento (60%), 200 exemplos para *Cross Validation* (20%) e 200 exemplos para conjunto de teste (20%). Tomamos dez funções hipótese, que são polinômios do primeiro até o décimo grau, e realizamos o treinamento, com cada uma delas.

1. $h_{\theta}(x) = \theta_0 + \theta_1 x$
2. $h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$
3. $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_3 x^3$
- \vdots
10. $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{10} x^{10}$

Em seguida, ao invés de verificarmos o erro do algoritmo diretamente no conjunto teste, comparamos primeiro as previsões de cada função hipótese com os dados do conjunto de *Cross Validation*, e selecionamos qual função hipótese apresenta menor erro. Dessa forma, verificamos o potencial da função hipótese de generalizar as previsões.

5 escoamento bifásico

Nesta seção apresentaremos brevemente o que adotaremos por sistema bifásico, alguns padrões de escoamento que ocorrem no escoamento bifásico e a definição de fração volumétrica.

O escoamento bifásico é caracterizado pelo escoamento simultâneo de duas fases que não se misturam. Neste escoamento podemos ter as fases em diferentes estados da matéria, como por exemplo: gás-líquido, líquido-líquido, gás-sólido e líquido sólido.

5.0.1 Padrão de escoamento

Em um escoamento bifásico, segundo [Mota \(2015\)](#), dois líquidos imiscíveis que possuem diferentes densidades e viscosidades, podem formar diversas configurações durante o escoamento dentro de uma tubulação. Essas configurações são denominadas *padrões de escoamento*.

Dentre os diversos padrões de escoamento, abordaremos três: o padrão anular, o padrão estratificado e o padrão mistura. Para exemplificação destes padrões, adotaremos sempre um sistema de escoamento bifásico líquido-líquido, formado por água e óleo.

O padrão anular é caracterizado por um círculo de escoamento de óleo, e o restante do escoamento de água, como é mostrado na [Figura 9](#):

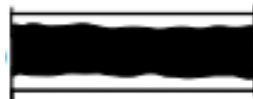


Figura 9 – Exemplo de padrão de escoamento anular, em vista axial.

Já no padrão estratificado, o escoamento da água e óleo é separado apenas por uma linha horizontal, sendo a fase superior a essa linha, o escoamento de óleo devido a sua menor densidade, e a fase inferior a de água. Podemos ver um exemplo deste escoamento, em vista axial na [Figura 10](#):



Figura 10 – Exemplo de padrão de escoamento estratificado, com vista axial.

O padrão mistura é caracterizado por vários pequenos círculos de escoamento de óleo cercados pela água, que se formam devido a uma maior turbulência durante o escoamento dos líquidos. Exemplo deste escoamento pode ser visto na [Figura 11](#).



Figura 11 – Exemplo de padrão de escoamento mistura com vista axial.

5.0.2 Fração Volumétrica

Fração volumétrica de uma fase é a razão entre o volume de uma fase específica e o volume total da mistura sendo escoada. Por exemplo, se adotarmos um dos padrões de escoamento apresentados anteriormente, a fração volumétrica de óleo é o percentual de óleo no escoamento, em relação ao líquido total sendo escoado. Esta fração é calculada considerando uma vista radial do escoamento, e portanto varia ao longo do tubo e ao longo do tempo.

6 Aplicação

Segundo [Mota \(2011\)](#), durante o processo de extração de petróleo, normalmente não se extraem somente os hidrocarbonetos desejados (gás e óleo), mas sim uma mistura de gás, óleo, água e impurezas. Para a obtenção dos produtos finais desejados, faz-se necessário o uso de equipamentos que realizem o processamento da mistura extraída.

Duas etapas importantes desse processamento são a separação de fases da mistura, e a medição volumétrica de óleo, gás e água após a separação, isto é, após serem separados gases, líquidos e sólidos, é preciso saber o volume que se está escoando dos gases e líquidos.

A medição das múltiplas fases no processo de escoamento apresenta respostas com representatividade significativas, quase em tempo real e reduz o número de recursos utilizados, o que traz economia para o procedimento de exploração de petróleo, principalmente em regiões de difícil acesso.

Para a medição multifásica no processo de escoamento de petróleo, um dos métodos que pode ser utilizado é a Tomografia por Capacitância Elétrica, que é similar a Tomografia por Impedância Elétrica. Neste caso, consideramos a utilização do modelo completo de eletrodos, aplicados a um tubo de escoamento de petróleo, para que dessa forma, obtenhamos um monitoramento do que está sendo escoado pelo tubo.

Nesse trabalho estudaremos a aplicação da Tomografia por Impedância Elétrica para o monitoramento, em tempo real, de escoamento bifásico (água-óleo) dentro de tubos de extração de petróleo.

A reconstrução de uma imagem que mostre exatamente a distribuição de água e óleo dentro do tubo em tempo real é praticamente impossível, visto que o TIE é um problema altamente mal-posto, conforme discutido anteriormente. Por esse motivo, propomos o uso de técnicas de ML para a identificação apenas do padrão de escoamento e da fração volumétrica de óleo.

Para que essa identificação seja feita, são lidos os potenciais elétricos nos eletrodos após a aplicação das correntes elétricas. A seguir, estes potenciais são aplicados no algoritmo (já treinado) de regressão linear, para determinação de fração volumétrica e na de regressão logística, para classificação do padrão de escoamento.

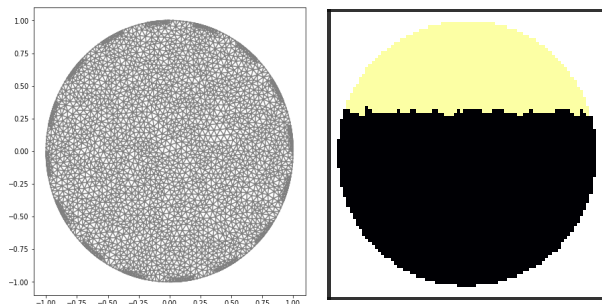
7 Experimentos Numéricos

Nesta seção apresentamos a aplicação das técnicas de Regressão Linear e de Regressão Logística em um conjunto de dados simulados do modelo completo de eletrodos de Tomografia por Impedância Elétrica aplicada a tubos de extração de petróleo.

Para a aplicação das técnicas de *Machine Learning*, utilizamos como base de dados três conjuntos de 1000 exemplos, um para cada padrão de escoamento apresentado na Seção 6.2, a saber, o padrão estratificado, o padrão anular e o padrão mistura. Cada exemplo é composto de um par de vetores que contém, respectivamente, dados de entrada e de saída, os quais serão detalhados mais adiante.

Os dados simulados foram gerados sob um círculo de raio 1 u.m. (unidade de medida), que denotamos por Ω , e imita uma seção radial de um tubo de escoamento de petróleo. No interior de Ω foi definida uma malha triangular com 5198 triângulos. As condutividades $\gamma : \Omega \rightarrow \mathbb{R}$ serão sempre funções constantes por partes, que no caso dos padrões de escoamento estratificado e anular tomam apenas os valores 1 e 2 (supomos por simplificação que essas são as condutividades do óleo e da água, respectivamente). No caso do padrão de escoamento estratificado os triângulos tomam valores 2 até uma determinada altura, e o restante dos triângulos tomam o valor 1. No caso do padrão de escoamento anular, os triângulos tomam o valor 2 dentro de uma circunferência e 1 fora dela, esta circunferência está localizada no interior do tubo e seu centro e raio variam.

Já no padrão de escoamento mistura, cada função γ é constante e toma valor entre 1 e 2, representando a condutividade da mistura dos dois líquidos. Uma imagem da malha e uma função γ representando um exemplo do padrão de escoamento estratificado, podem ser vistos na Figura 12, em que a parte em amarelo representa a condutividade tomando valor 1 (óleo) e a parte preta representa a condutividade 2 (água).



(a) Malha de Ω

(b) Função γ representando um exemplo do padrão de escoamento estratificado

Figura 12 – Exemplo de padrão de escoamento estratificado.

Na fronteira de Ω , que denotaremos por $\partial\Omega$, foram acoplados 16 eletrodos, denotados por e_l , $l = 1, 2, \dots, 16$, de tamanhos iguais, igualmente espaçados, e cobrindo metade de $\partial\Omega$.

Foram fixadas 16 correntes elétricas, que denotamos por I^j , $j = 1, 2, \dots, 16$, com $I^j \in \mathbb{R}^{16}$. Cada uma dessas correntes é representada por uma linha da matriz $I \in \mathbb{R}^{16 \times 16}$, apresentada abaixo. Podemos escrever também $I = (I^1, I^2, \dots, I^{16})$.

$$I = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

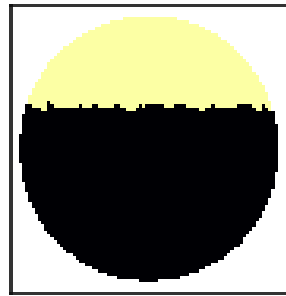
Para cada exemplo de condutividade γ aplicam-se as 16 correntes elétricas I^j , e são mensurados os potenciais gerados, que denotamos por U^j , $j = 1, 2, \dots, 16$, com $U^j \in \mathbb{R}^{16}$. Diferentemente das correntes, organizamos os potenciais elétricos em forma de vetor $U \in \mathbb{R}^{256}$, com $U = (U^1, U^2, \dots, U^{16})$.

Assim, como cada um dos três padrões de escoamento possui 1000 exemplos, dispomos os potenciais elétricos em matrizes de dimensão 1000×256 , em que a i -ésima linha representa os potenciais elétricos associados ao i -ésimo exemplo de condutividade $\gamma^{(i)}$. As linhas dessas matrizes representarão os vetores $x^{(i)}$, com $i = 1, 2, \dots, 1000$, que serão as entradas para o problema de *Machine Learning*.

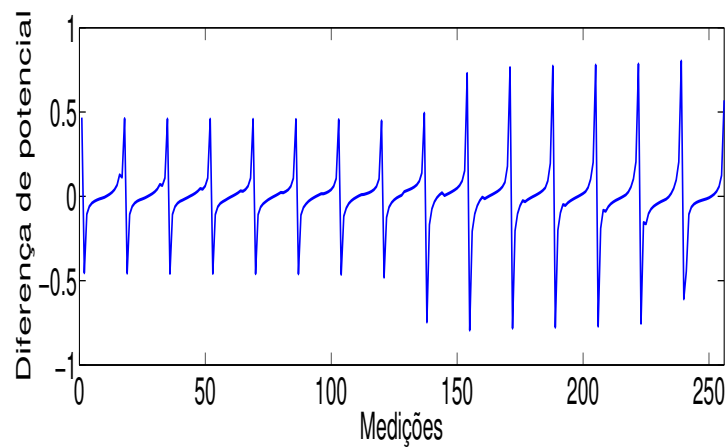
As saídas $y^{(i)}$ serão de dois tipos diferentes. Na seção 8.1, estudaremos o problema de regressão linear, e os vetores $y^{(i)} \in [0, 1]$, representarão a fração volumétrica de óleo referente ao i -ésimo exemplo. Perceba que como $\gamma^{(i)}$ assume valor 1 no óleo e 2 na água, tem-se que $y^{(i)} = 1 - \frac{1}{\pi} \int_{\Omega} \gamma^{(i)}$. Já na seção 8.2, estudaremos o problema de regressão logística, e nesse caso gostaríamos de classificar os padrões de escoamento, por isso os vetores $y^{(i)} \in \{0, 1\}$. Os detalhes serão explicados nas seções mencionadas.

Portanto, os problemas de *Machine Learning* serão tratados utilizando os pares $(x^{(i)}, y^{(i)}) \in \mathbb{R}^{256} \times \mathbb{R}$. Ainda, para facilitar a operação entre matrizes sem causar qualquer prejuízo aos algoritmos, adicionaremos uma entrada de valor unitário aos vetores $x^{(i)}$, de forma que $x^{(i)} = [1, x_1^{(i)}, x_2^{(i)}, \dots, x_{255}^{(i)}, x_{256}^{(i)}]$, e portanto $x^{(i)} \in \mathbb{R}^{257}$.

Agora, para exemplificar, vamos mostrar na Figura 13, uma função $\gamma^{(i)}$ do padrão de escoamento estratificado e os potenciais elétricos associados. Neste caso, os potenciais correspondem a um vetor $x^{(i)}$, o qual deve ser utilizado para determinar a fração volumétrica ou o padrão de escoamento da função $\gamma^{(i)}$.



(a) Exemplo de condutividade $\gamma^{(i)}$ do padrão de escoamento estratificado



(b) Diferenças de Potenciais elétricos associados a condutividade $\gamma^{(i)}$

Figura 13 – Exemplo de condutividade elétrica com o potencial correspondente.

Os algoritmos de *Machine Learning* foram implementados com o auxílio do *software* MatLab R2013a. Para a implementação dos algoritmos, utilizamos 60% dos dados para treinamento, 20% para o conjunto de *Cross Validation* e os 20% restantes deixamos para o conjunto de teste.

7.1 Determinação da Fração Volumétrica

Para a determinação da fração volumétrica treinamos algoritmos de *Machine Learning* com a técnica de Regressão Linear (RL), utilizando dois métodos distintos: (a) um método do tipo gradiente, chamado Método de Landweber, que será abordado na subseção 8.1.1; (b) o método de Tikhonov, que será abordado na subseção 8.1.2.

Reforçamos que o intuito é que o algoritmo determine a fração volumétrica no interior de um tubo de escoamento, com base nas diferenças de potencial obtidas a partir da configuração de condutividade em seu interior, após a aplicação das correntes elétricas.

Cada padrão de escoamento será tratado separadamente. Isso implica que será feita uma regressão linear utilizando o método de Landweber e outra utilizando o método de Tikhonov para cada padrão. Lembre que cada padrão de escoamento está associado a uma matriz pertencente a $\mathbb{R}^{1000 \times 257}$, em que a i -ésima linha representa o vetor $x^{(i)}$, e $y^{(i)}$ representa a fração volumétrica do i -ésimo exemplo.

Para ilustrar, na Figura 14 temos dois exemplos de condutividade do padrão de escoamento estratificado e os potenciais correspondentes.

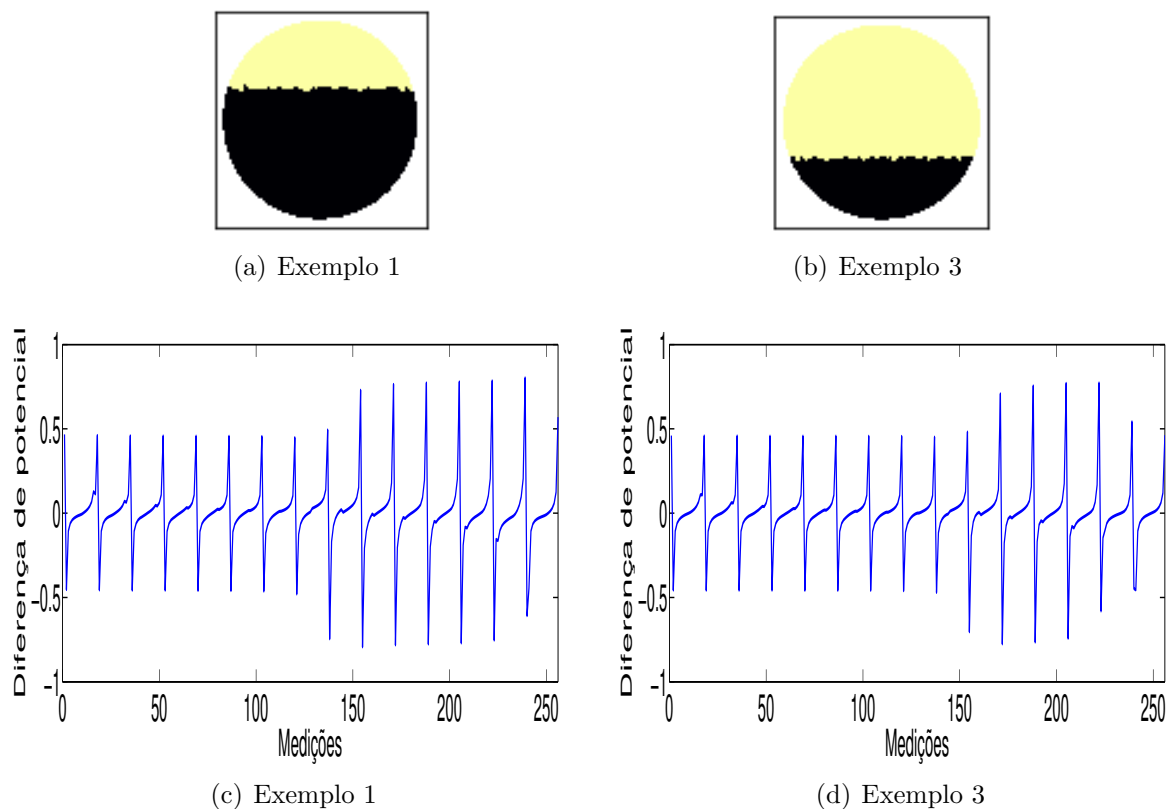


Figura 14 – Exemplos de condutividade e respectivos potenciais elétricos.

Podemos ver que a diferença de fração volumétrica é nítida, enquanto a diferença entre os potenciais elétricos não apresenta a mesma característica. Isto é uma característica de problemas mal-postos.

Com o intuito de evitar que o algoritmo seja tendencioso, em cada conjunto fizemos uma permutação dos 1000 exemplos, através do *MatLab*. Em seguida aplicamos a mesma permutação aos dados de saída, para que as matrizes não percam a correspondência de dados. Na sequência separamos os 600 primeiros exemplos para o conjunto de treinamento, os 200 exemplos seguintes para o conjunto de *Cross Validation*, e os 200 restantes para o conjunto teste.

Como função hipótese da regressão linear, utilizamos a seguinte função:

$$h_{\theta}(x^{(i)}) = \theta_0 1 + \theta_1 x_1^{(i)} + \dots + \theta_{255} x_{255}^{(i)} + \theta_{256} x_{256}^{(i)} = \langle \theta, x^{(i)} \rangle \quad (7.1)$$

A respectiva função custo utilizada para o treinamento, conforme (7.3) é:

$$J(\theta) = \frac{1}{1200} \sum_{i=1}^{600} (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (7.2)$$

Podemos reescrever a função custo na forma vetorial:

$$J(\theta) = \frac{1}{1200} \|X\theta - Y\|^2 \quad (7.3)$$

com $X \in \mathbb{R}^{600 \times 257}$, $\theta \in \mathbb{R}^{257}$ e $Y \in \mathbb{R}^{600}$, dados por:

$$X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(600)} \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{256} \end{bmatrix}, \quad Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(600)} \end{bmatrix}.$$

7.1.1 Método de Landweber

Nessa subseção vamos aplicar a técnica de regressão linear utilizando o método de Landweber, para cada um dos três padrões de escoamento. Vamos apresentar o procedimento detalhado do treinamento, CV e teste para o padrão mistura. Para os demais padrões apresentaremos apenas o resultado final do conjunto teste.

Na execução do método de Landweber, partimos em todos os casos do vetor nulo, isto é, $\hat{\theta}_0 = 0$, $\hat{\theta} \in \mathbb{R}^{257}$. Em cada passo, seguimos na direção oposta do gradiente do funcional (7.3) resultando na seguinte iteração:

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \lambda \nabla J(\hat{\theta}_n) \quad (7.4)$$

No nosso caso, a equação (7.4) fica da forma:

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \frac{\lambda}{600} X^T (X\hat{\theta}_n - Y) \quad (7.5)$$

conforme (2.6) e (2.7).

Para o treinamento da regressão linear, foram executadas 10000 iterações. O tamanho de passo λ foi determinado por tentativa e erro, e após várias tentativas, o maior tamanho de passo λ encontrado, sem que o algoritmo divergisse foi $\lambda = 0,1$. Podemos conferir a evolução da sequência $J(\hat{\theta}_k)$ na Figura 15.

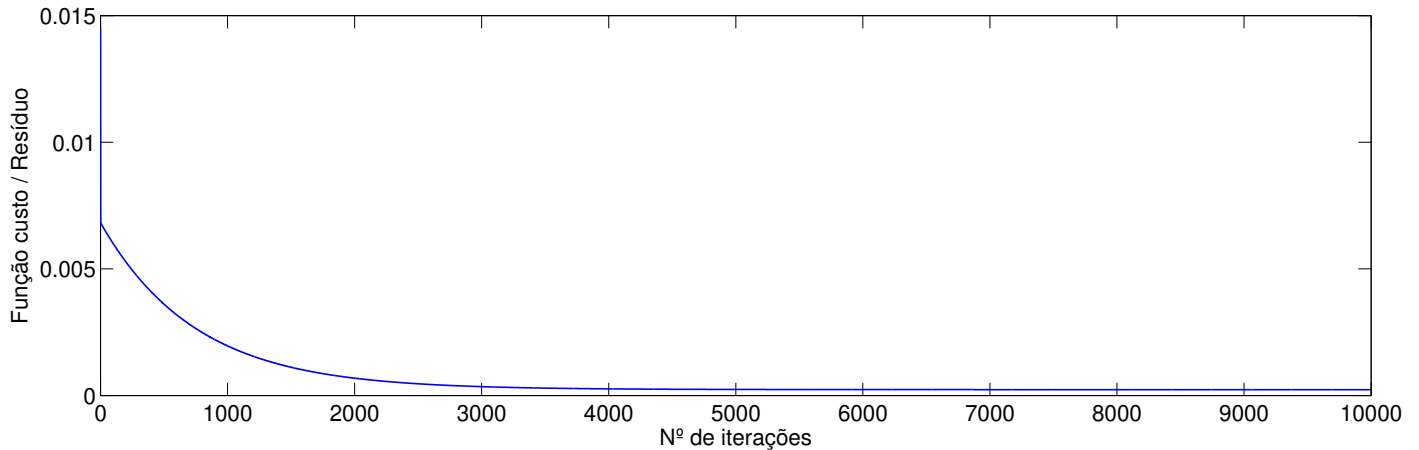


Figura 15 – Valor da função custo conforme cada iteração no treinamento do algoritmo de regressão linear.

Para realizar o CV, fizemos um procedimento diferente do explanado na Seção 5. Utilizamos os 10000 vetores $\hat{\theta}_k$ obtidos do treinamento para obter os resíduos:

$$a_k = \frac{1}{200} \|\hat{\theta}_k X - Y\|^2 \quad (7.6)$$

em que $X \in \mathbb{R}^{200 \times 257}$ e $Y \in \mathbb{R}^{200}$, formam o conjunto de CV. A seguir selecionamos o índice $k = \hat{k}$, em que a_k atinge seu valor mínimo.

Na figura 16 apresentamos a evolução da sequência a_k . A iteração \hat{k} obtida foi a última, isto é, $\hat{k} = 10000$.

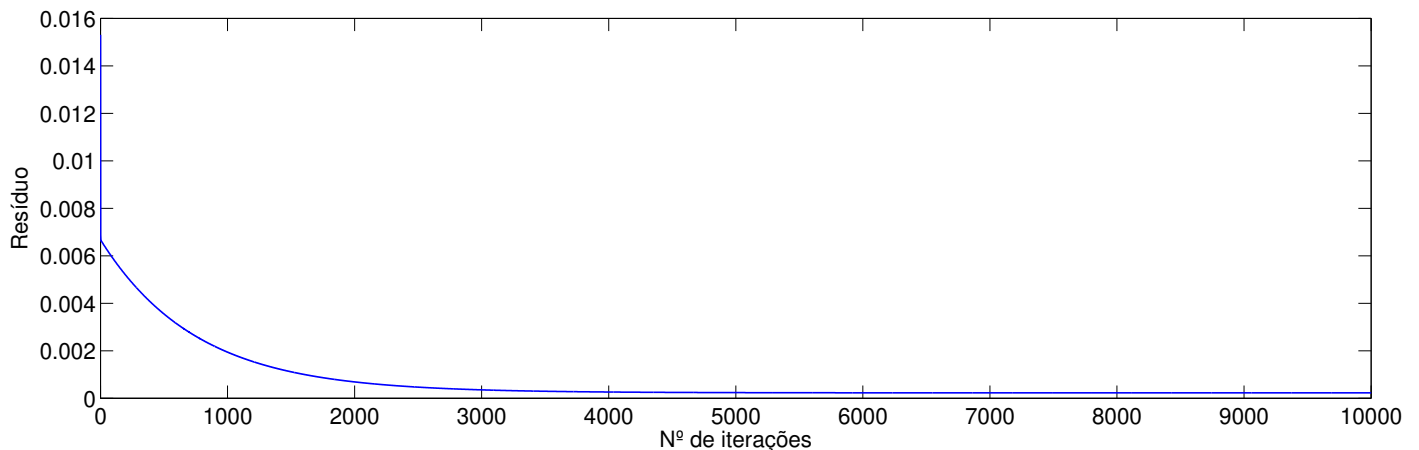


Figura 16 – Evolução do resíduo no conjunto de *Cross Validation*.

Observe pela Figura 16 que o resíduo no conjunto de CV reduz conforme o índice k aumenta, da mesma forma que ocorre no conjunto de treinamento (Figura 15). A partir disso, podemos verificar que não ocorre *underfitting*, pois o resíduo no conjunto de treinamento reduziu até o valor do erro médio de 0,00023, isto mostra que a função hipótese é complexa o suficiente para o nosso propósito. Por outro lado também não ocorre *overfitting*, pois no conjunto de CV, mesmo com dados não vistos pelo algoritmo, o resíduo também diminuiu a um valor do erro médio de 0,00022.

Cabe ressaltar que, não era esperado que o resíduo no conjunto de CV fosse menor do que o resíduo no conjunto de treinamento, dado que no conjunto CV haviam dados não vistos pelo algoritmo. Entretanto, como ambos apresentaram erros pequenos, seguimos com o procedimento para o conjunto teste.

Por último, selecionamos o vetor que apresentou menor resíduo no conjunto de CV ($\hat{\theta}_{10000}$) e aplicamos a todos os 200 vetores $x^{(i)}$ do conjunto teste, para prever os valores das frações volumétricas $y^{(i)}$. Podemos observar os resultados na Figura 17. Para facilitar a visualização, fizemos uma permutação dos vetores das frações volumétricas $y^{(i)}$ exatas, para que ficassem em ordem crescente, e aplicamos a mesma permutação nas previsões.

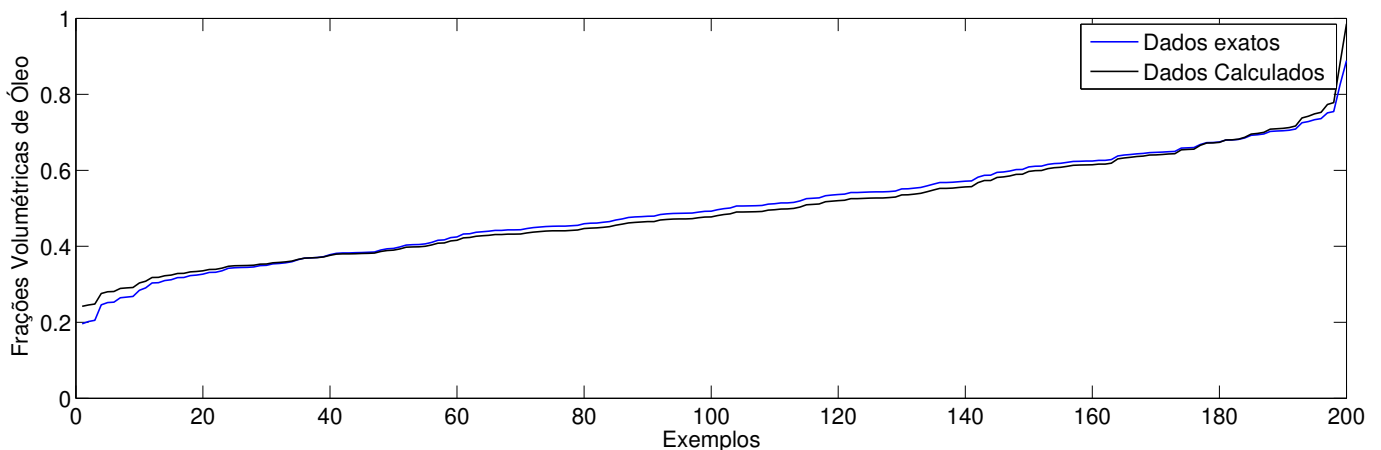
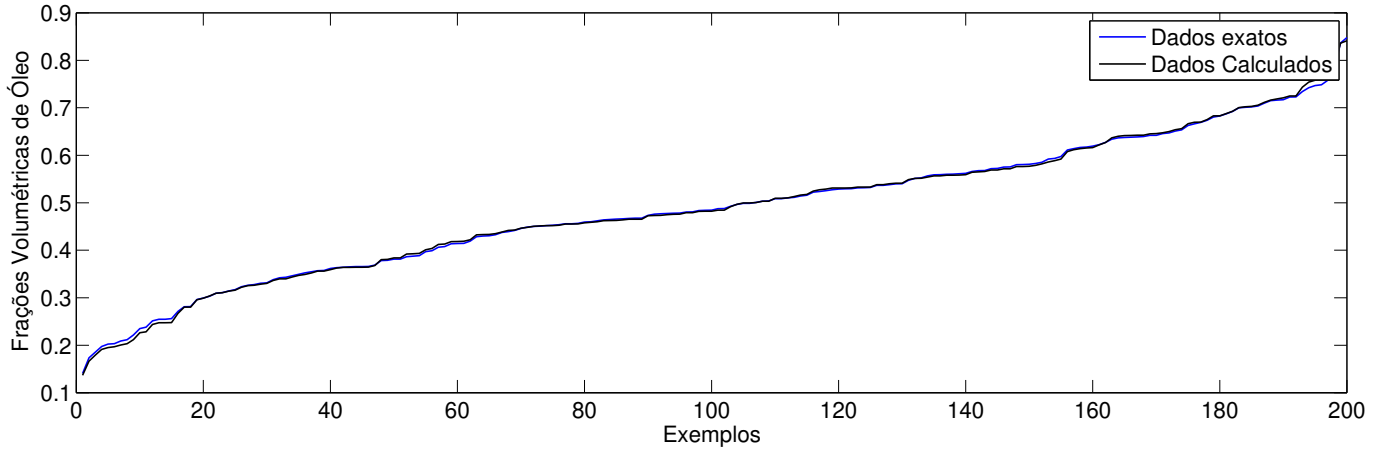


Figura 17 – Conjunto teste: Comparação entre as frações volumétricas exatas $y^{(i)}$ e as calculadas $\langle \hat{\theta}_{10000}, x^{(i)} \rangle$.

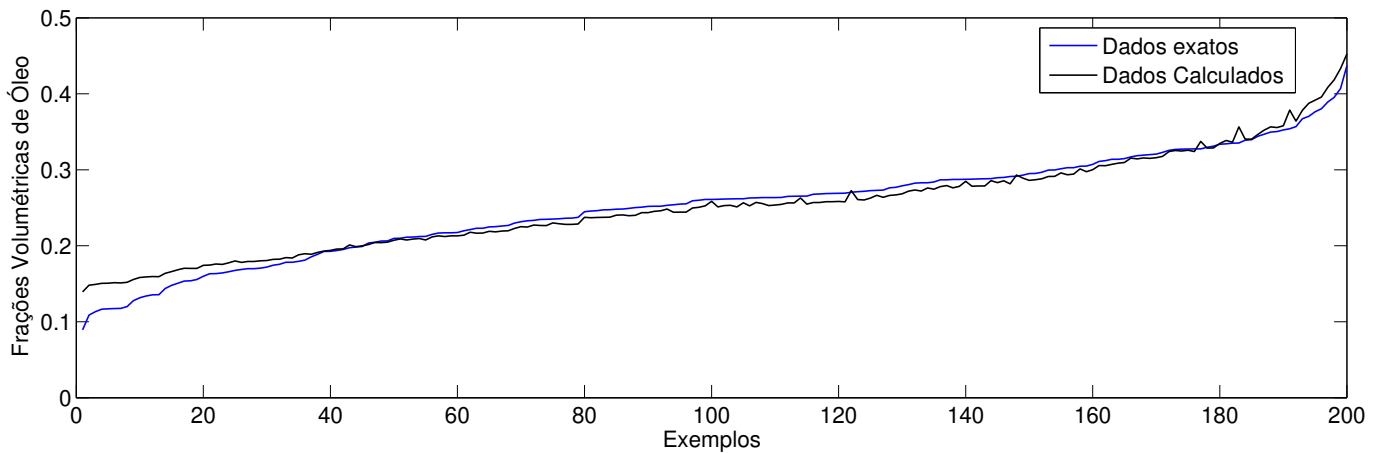
De forma análoga, fizemos o treinamento para o padrão de escoamento estratificado e padrão de escoamento anular, obtendo os tamanhos de passo para o método de Landweber de $\lambda = 0,1$ e $\lambda = 0,2$, respectivamente. Em ambos os padrões também realizamos 10000 iterações.

Tanto no padrão de escoamento estratificado, quanto no padrão de escoamento anular, o comportamento da função custo no conjunto de treinamento, bem como o resíduo no conjunto de CV, apresentaram resultados muito semelhantes aos vistos no treinamento e CV do padrão de escoamento mistura.

Devido a similaridade dos comportamentos, apresentaremos apenas a comparação dos vetores selecionados após o CV com os vetores exatos dos conjuntos teste.



(a) Padrão de escoamento estratificado



(b) Padrão de escoamento anular

Figura 18 – Conjunto teste: comparação das frações volumétricas exatas e calculadas.

7.1.2 Método de Tikhonov

Nesta subseção também vamos aplicar a técnica de regressão linear, porém agora utilizaremos o método de Tikhonov, para cada um dos três padrões de escoamento. Para fim de comparação com o método da subseção anterior, também apresentaremos o procedimento detalhado do treinamento, CV e teste para o padrão mistura.

No método de Tikhonov, substituímos a função custo (7.3), pelo funcional de Tikhonov associado a função custo, que é:

$$T_{\alpha_k}(\hat{\theta}) = \frac{1}{2} \|X\hat{\theta} - Y\|^2 + \frac{\alpha_k}{2} \|\hat{\theta}\|^2 \quad (7.7)$$

Em seguida, para encontrar o minimizador do funcional de Tikhonov, resolvemos a equação $\nabla T_{\alpha_k}(\hat{\theta}) = 0$, que resulta em:

$$\hat{\theta}_k = (X^T X + \alpha_k I)^{-1} X^T Y \quad (7.8)$$

Para realizarmos o treinamento, resolvemos a equação (7.8) com os parâmetros de regularização $\alpha_k = 1000(0,8)^k$, $k = 0, 1, \dots, 40$, e plotamos $J(\hat{\theta}_k)$ vs. k , em que J está definida em (7.3), como pode ser observado na Figura 19.

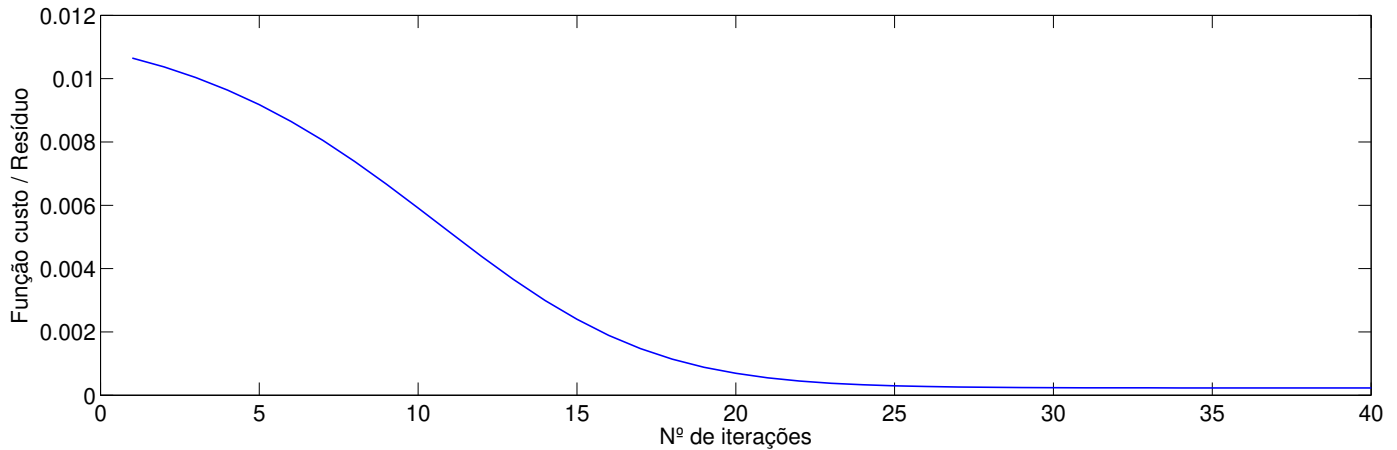


Figura 19 – Valor da função custo conforme cada iteração no treinamento para o padrão mistura.

Em seguida, realizamos um procedimento similar ao da subseção anterior para executar o processo de CV. Utilizamos os 40 vetores obtidos do treinamento, e calculamos os resíduos:

$$b_k = \frac{1}{200} \|X\hat{\theta}_k - Y\|^2 \quad (7.9)$$

em que $X \in \mathbb{R}^{257 \times 200}$ é a matriz de potenciais do conjunto de CV e $Y \in \mathbb{R}^{200}$ é o vetor de frações volumétricas do conjunto CV. Em seguida, selecionamos o índice $k = \hat{k}$, em que b_k apresenta o menor valor.

Podemos acompanhar a evolução da sequência b_k na Figura 20. O índice escolhido foi $\hat{k} = 40$.

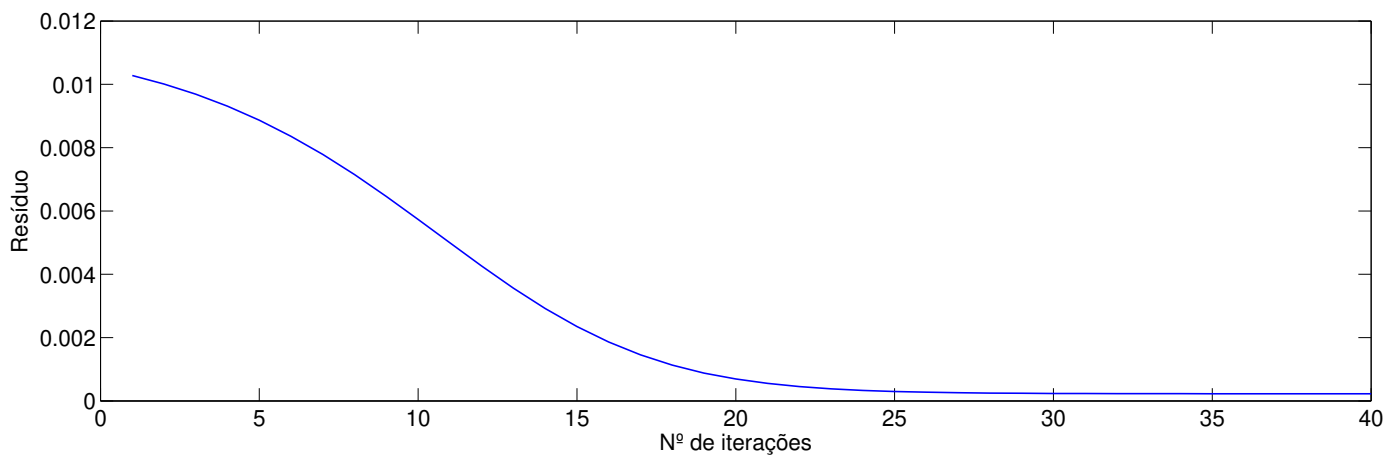


Figura 20 – Resíduo no conjunto de *Cross Validation* no padrão de escoamento mistura.

Novamente, podemos verificar que não ocorrem *underfitting* e *overfitting*, pela mesma argumentação feita na subseção anterior. Os resíduos atingidos na iteração $\hat{\theta}_{40}$ no conjunto de treinamento e CV foram 0,00023 e 0,00022, respectivamente.

A comparação das frações calculadas usando o vetor $\hat{\theta}_k$, com o vetor de frações volumétricas exatas do conjunto teste, pode ser observada na Figura 21.

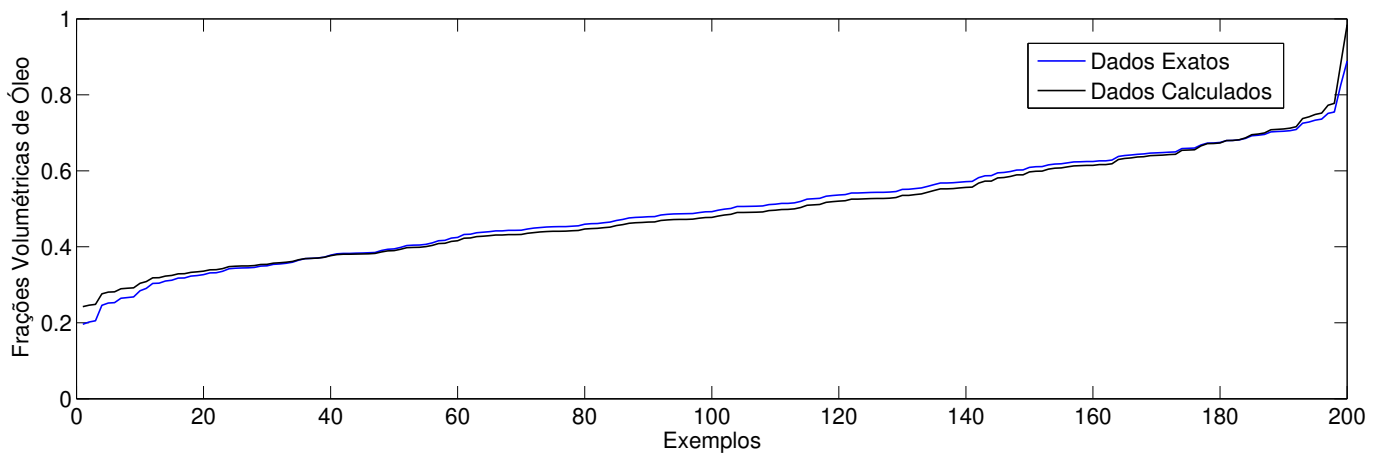


Figura 21 – Conjunto teste: Comparação das frações volumétricas exatas com as calculadas.

7.1.2.1 Comparação entre Método de Landweber e Método de Tikhonov

Separamos esta subseção para compararmos o desempenho dos dois métodos utilizados para determinação da fração volumétrica.

No caso do padrão de escoamento mistura, podemos verificar o gráfico comparativo na figura:

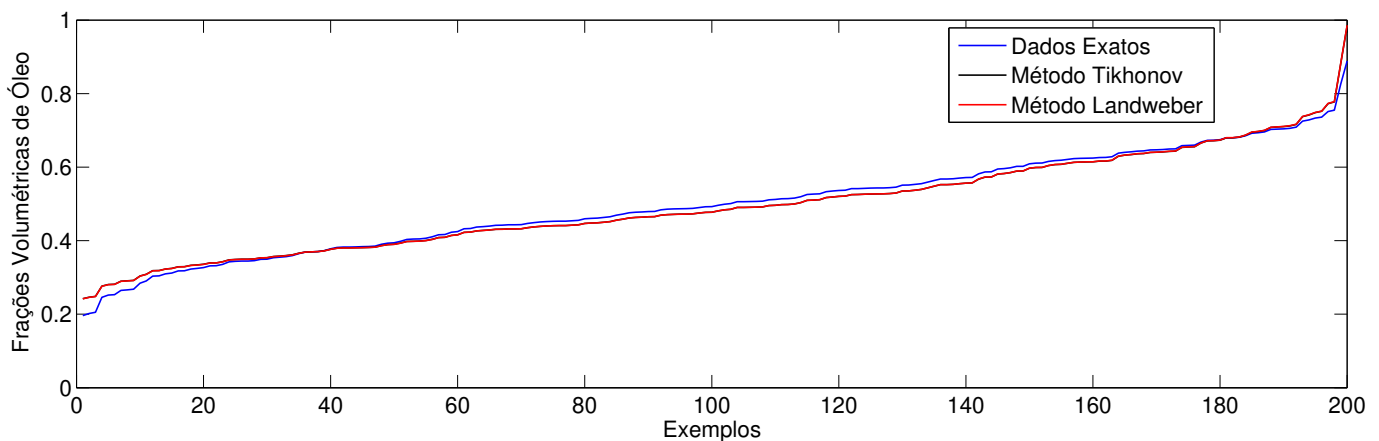


Figura 22 – Comparação do desempenho dos métodos de Tikhonov e Landweber.

Podemos perceber que os gráficos de ambos os métodos, praticamente se sobrepõem. Ao somarmos o módulo das diferenças das frações calculadas pelas frações exatas, exemplo a exemplo, e dividirmos pelo número de amostras (200 neste caso), obtemos tanto no Método de Landweber, quanto no método de Tikhonov um erro médio de 0,012.

Cabe salientar que a forma de resolução do sistema linear no método de Tikhonov é através da eliminação Gaussiana.

7.2 Determinação do padrão de escoamento

Na determinação entre os três padrões de escoamento, treinamos algoritmos com a técnica de Regressão Logística Múltipla, utilizando o Método de Landweber.

Nesta técnica a partir dos dados dos potenciais elétricos gerados, o algoritmo deve determinar a qual dos três padrões de escoamento o exemplo pertence: padrão mistura, estratificado ou anular.

Para ilustrar, temos na Figura 23, dois exemplos de condutividade γ em Ω , que representam dois padrões de escoamento distintos, e em seguida, a imagem de seus respectivos potenciais elétricos.

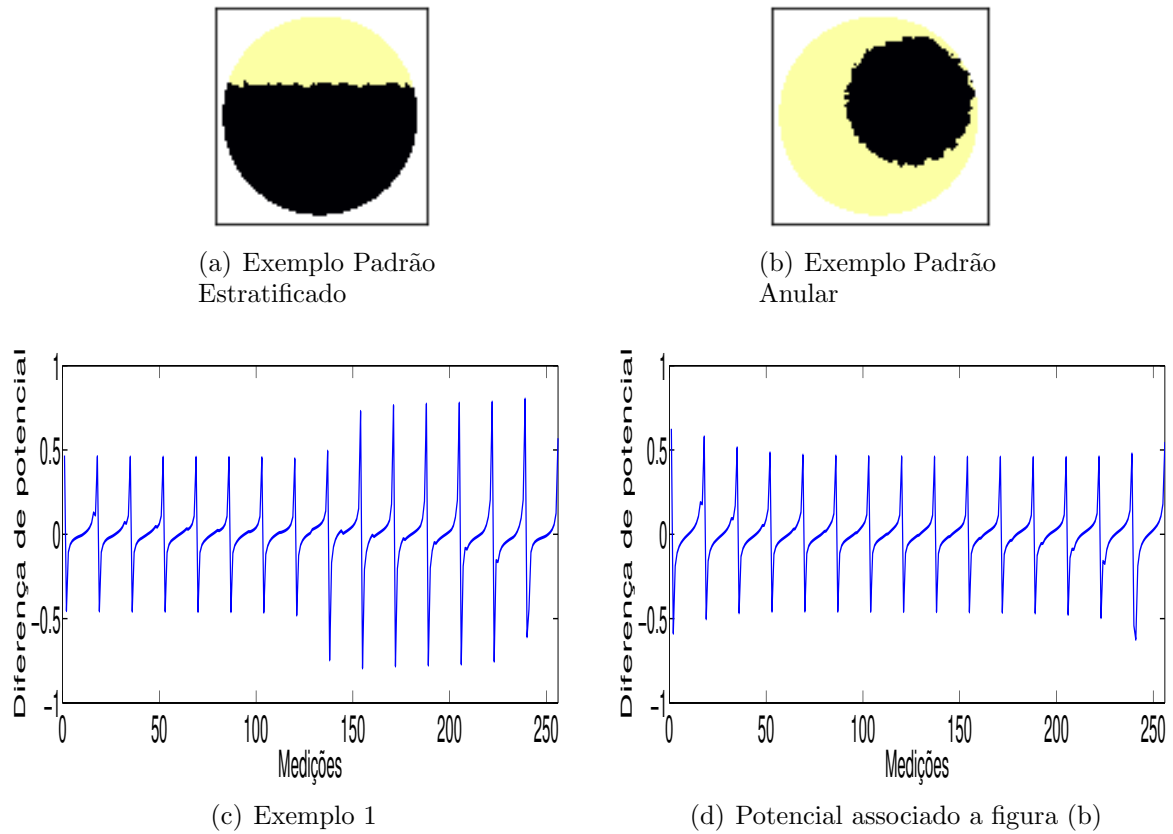


Figura 23 – Exemplos de diferentes padrões de escoamento e os respectivos potenciais.

Para a aplicação da técnica de Regressão Logística Múltipla, precisamos realizar três Regressões Logísticas Simples, uma para cada padrão de escoamento. Antes de realizar as regressões, com o intuito de evitar que o algoritmo seja tendencioso, embaralhamos as 1000 linhas da matriz de potenciais de cada padrão de escoamento, através de uma permutação de inteiros. Depois, tomamos as primeiras 600 linhas das matrizes do padrão mistura, do estratificado e do anular, e empilhamos nesta mesma ordem, uma sobre a outra numa mesma matriz. Esta nova matriz de dimensão $\mathbb{R}^{1800 \times 257}$ destinamos para o conjunto de treinamento, que representa 60% do total dos exemplos. Os 200 exemplos subsequentes de cada matriz embaralhada inicialmente, fazemos o mesmo procedimento de empilharmos os dados, na mesma ordem de padrões, obtendo uma matriz com 600 linhas, que forma o conjunto de *Cross Validation*. Os 200 restantes de cada matriz, repetimos o mesmo procedimento, e utilizamos para o conjunto de teste. Com os dados organizados, realizamos as regressões logísticas.

Como queremos classificar, não nos são necessárias as informações das frações volumétricas. Assim, para a regressão logística do padrão de escoamento mistura, por exemplo, associamos a matriz do conjunto treinamento um vetor $Y \in \mathbb{R}^{1800}$, em que as 600 primeiras entradas tomam o valor 1, e as 1200 entradas restantes tomam o valor 0. Dessa forma, o valor 1 fica associado aos exemplos do padrão mistura, e o valor 0 para os demais.

Para o padrão de escoamento estratificado e anular, realizamos a mesma associação da matriz de treinamento com vetores $Y \in \mathbb{R}^{1800}$, mas cada vetor Y possui o valor 1 nas entradas que correspondem aos exemplos do padrão que se está realizando a técnica de regressão linear.

Vamos apresentar agora o treinamento das regressões logísticas através do método de Landweber, e o respectivo CV, para cada padrão de escoamento.

A função hipótese para cada regressão linear, conforme (4.3), é:

$$h_{\theta}(X) = g(X\theta) \quad (7.10)$$

em que:

$$g(z) = \frac{1}{1 + e^{-z}} \quad (7.11)$$

Neste caso, cabe ressaltar que $h_{\theta} : \mathbb{R}^{1800 \times 257} \rightarrow \mathbb{R}^{1800}$, isto é, a função sigmoide $g : \mathbb{R}^{1800} \rightarrow \mathbb{R}^{1800}$ é aplicada em cada entrada do produto $X\theta$.

A função custo conforme (4.5), é:

$$J(\theta) = \frac{1}{1800} [-Y^T \log(h_{\theta}(X)) - (1 - Y^T) \log(1 - h_{\theta}(X))] \quad (7.12)$$

Da mesma forma que a função sigmoide, a função log é calculada entrada a entrada.

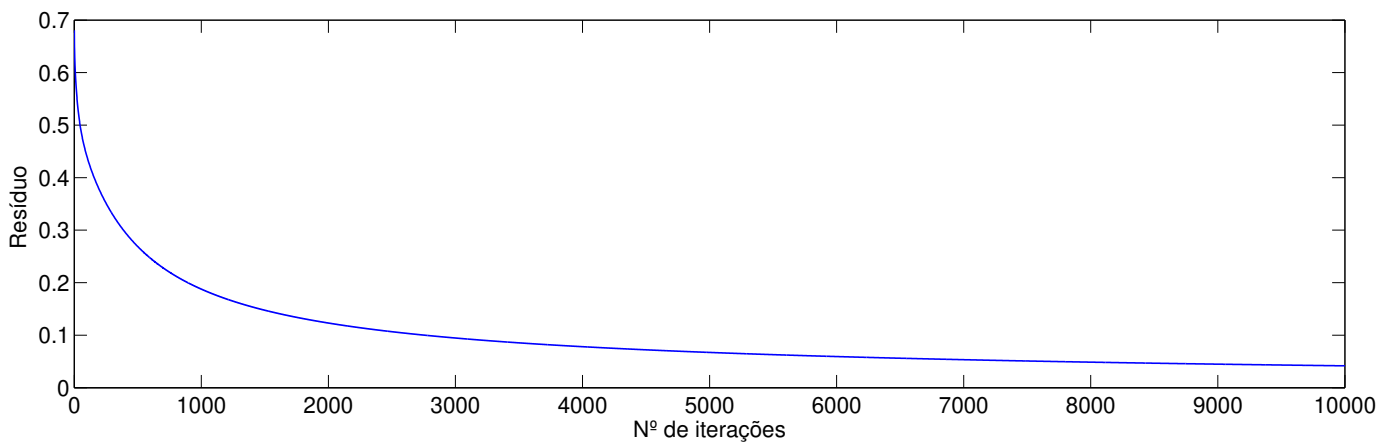
De forma similar a determinação da fração volumétrica, para o treinamento das regressões logísticas através do método de Landweber, partimos do vetor nulo $\hat{\theta}_0 = 0$, e na iteração $k + 1$ seguimos na direção oposta ao vetor gradiente de $J(\hat{\theta}_k)$. As iterações ficam da seguinte forma:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \frac{\lambda}{1800} X^T (h_{\hat{\theta}_k}(X) - Y) \quad (7.13)$$

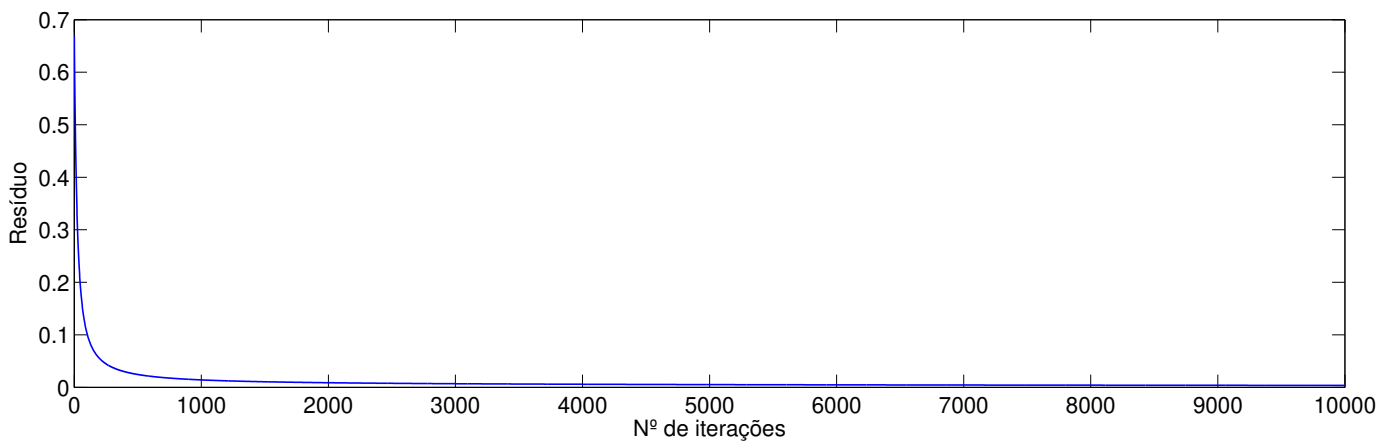
já que:

$$\nabla J(\theta) = \frac{1}{1800} X^T (h_{\theta}(X) - Y) \quad (7.14)$$

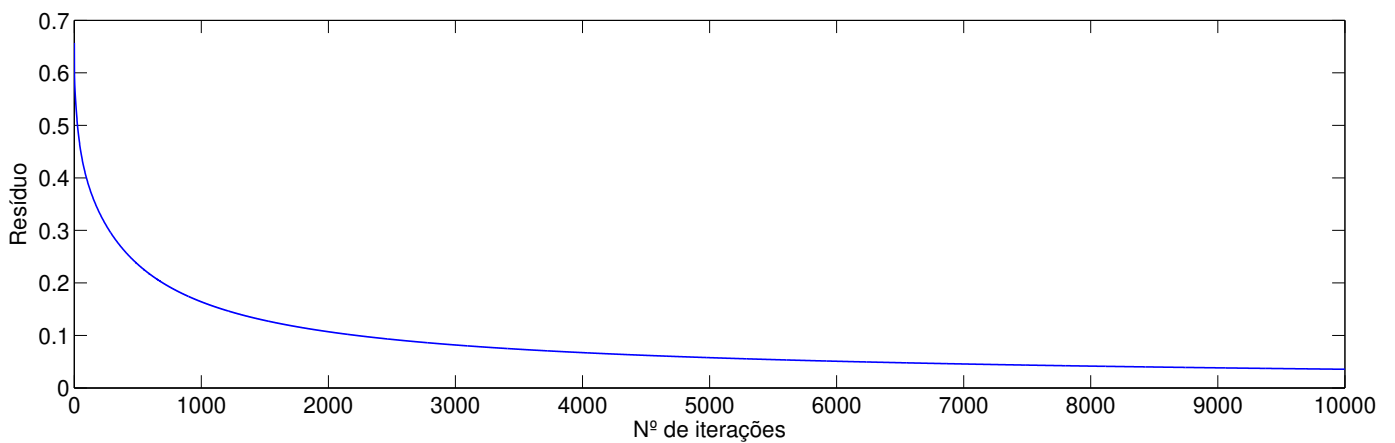
Para o treinamento das funções hipótese, realizamos 10000 iterações para cada padrão de escoamento. O tamanho de passo obtido por tentativa e erro, foi de $\lambda = 0,7$ para todos os três padrões de escoamento. Podemos ver nas imagens da Figura 24, a evolução das funções custo $J(\hat{\theta}_k)$ para o treinamento da regressão logística, em cada um dos padrões de escoamento.



(a) Treinamento padrão Mistura



(b) Treinamento padrão Estratificado



(c) Treinamento padrão Anular

Figura 24 – Treinamento das regressões logísticas para cada padrão de escoamento.

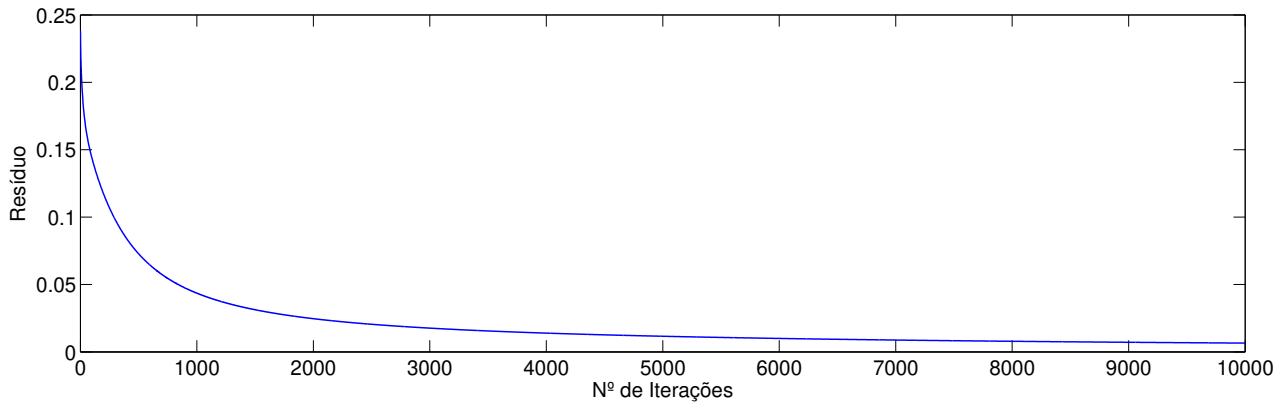
Podemos observar que o comportamento dos três gráficos é similar, com a distinção de que o valor da função custo no treinamento para o padrão estratificado decai mais rapidamente.

Para realizarmos o CV, usamos a matriz X e o vetor Y do CV para computarmos a

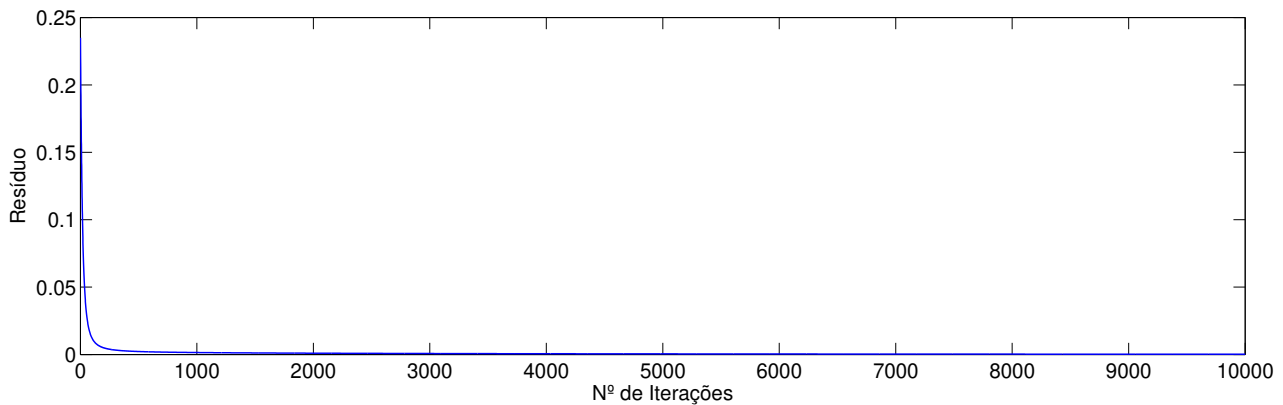
seguinte sequência:

$$c_k = \frac{1}{200} \|h_{\hat{\theta}_k}(X) - Y\|^2 \quad (7.15)$$

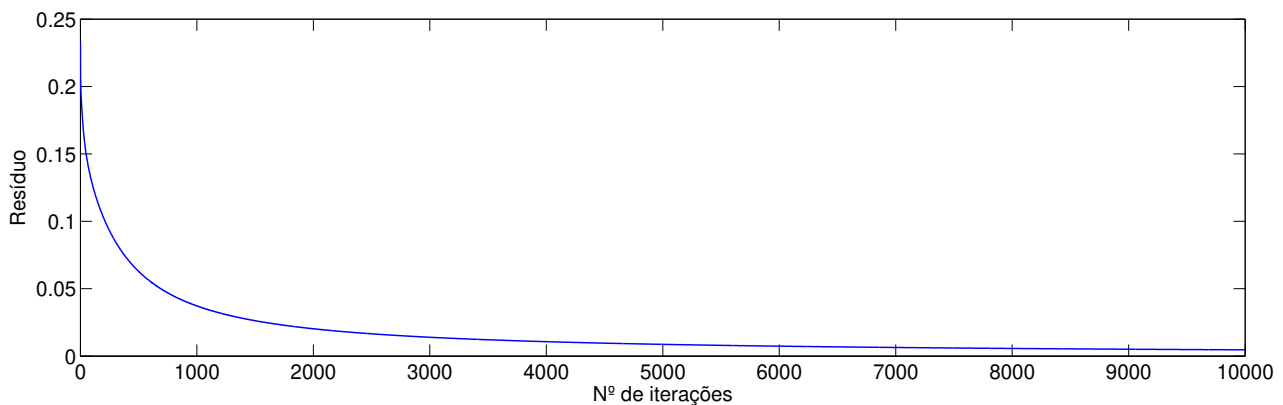
Podemos acompanhar a evolução da sequência c_k para cada padrão na Figura 25.



(a) Padrão Mistura



(b) Padrão Estratificado



(c) Padrão Anular

Figura 25 – Resíduo nos conjunto de *Cross Validation* das regressões logísticas para cada padrão de escoamento.

Nos três casos, a iteração que apresentou menor resíduo foi a última, isto é, $\hat{k} = 10000$.

Da mesma forma que foi argumentado na subseção 8.1, verifica-se que não há *overfitting* e *underfitting*.

Dando sequência, para executar a classificação no conjunto teste, tomamos os três vetores $\hat{\theta}_{10000}$, que foram os que apresentaram menor resíduo no conjunto CV. A partir de cada um desses vetores, aplicamos $h_{\hat{\theta}_{10000}}(X)$, em que $X \in \mathbb{R}^{600 \times 257}$ são os potenciais do conjunto teste. Assim, cada exemplo do conjunto teste está associado a três valores, um para cada função hipótese $h_{\hat{\theta}_{10000}}$, isto é, um para cada padrão de escoamento. O padrão cuja função hipótese retornar o maior valor dos três, é no qual o exemplo será classificado.

Podemos observar o resultado dessa classificação na Tabela 1. Nesta tabela, nas colunas estão os nomes dos padrões de escoamento que são as classificações exatas dos exemplos, já nas linhas, são as classificações feitas pelo algoritmo.

Calculados/ Exatos	Padrão Mistura	Padrão Estratificado	Padrão anular
Padrão Mistura	200	0	0
Padrão Estratificado	0	200	0
Padrão Anular	0	0	200

Tabela 1 – Resultado da classificação no conjunto teste.

Analisando a tabela, vemos que o algoritmo de regressão logística classificou corretamente todos os exemplos do conjunto teste, resultado este que permite avançarmos para o próximo passo, que é testar o algoritmo em dados com ruído.

7.3 Dados com ruído

Nesta subseção, repetimos os procedimentos dos algoritmos das subseções anteriores, tanto para a determinação de fração volumétrica quanto para a classificação do padrão de escoamento. Porém agora, iremos inserir artificialmente ruído nos dados de entrada, isto é, nos potenciais elétricos. Essa inserção de ruído ocorrerá apenas nos conjuntos de *Cross Validation* e teste, sendo que o conjunto de treinamento permanecerá sem ruído. Realizamos este procedimento pois sempre teremos ruído na mensuração de dados. O objetivo é verificar a performance do algoritmo na presença de dados com ruído.

O ruído é inserido individualmente em cada exemplo. Para exemplificar, seja $x^{(i)}$ o i -ésimo exemplo de um dos padrões de escoamento. Definimos o vetor com ruído, respectivo a $x^{(i)}$, com a porcentagem de ruído δ , da seguinte forma:

$$x_{\delta}^{(i)} = x^{(i)} + \delta \|x^{(i)}\| \frac{\eta^{(i)}}{\|\eta^{(i)}\|} \quad (7.16)$$

em que $\eta^{(i)} \in \mathbb{R}^{256}$ é um vetor aleatório gerado para cada exemplo, através do comando *rand* do *MatLab* que utiliza a distribuição uniforme, com $\eta^{(i)} = [\eta_1^{(i)}, \eta_2^{(i)}, \dots, \eta_{256}^{(i)}]$, e cada entrada $-1 \leq \eta_j^{(i)} \leq 1$, $j = 1, 2, \dots, 256$. Desse modo temos que $\frac{\|x_{\delta}^{(i)} - x^{(i)}\|}{\|x^{(i)}\|} = \delta$.

Após inserido o ruído, adicionamos uma entrada unitária extra como primeira entrada nos vetores $x_\delta^{(i)}$, como já mencionado, para facilitar a operação matricial, de forma que agora $x_\delta^{(i)} \in \mathbb{R}^{257}$.

Faremos a aplicação para os valores de δ de 0,1, 0,3 e 0,5, isto é, 10%, 30% e 50% de nível de ruído. Cabe ressaltar que os ruídos não são cumulativos, isto é, após inserido o ruído de 10% por exemplo, realizamos os procedimentos no conjunto CV e teste até finalizar os experimentos. Para testar com o próximo nível de ruído, iniciamos novamente o procedimento de inserção de ruído, a partir dos dados originais.

7.3.1 Determinação de Fração Volumétrica

Na determinação da fração volumétrica, apresentaremos apenas os procedimentos feitos com o Método de Landweber.

Para retomar o raciocínio e ilustrar, apresentamos na Figura 26, o gráfico da função custo no treinamento da regressão linear para o padrão de escoamento estratificado.

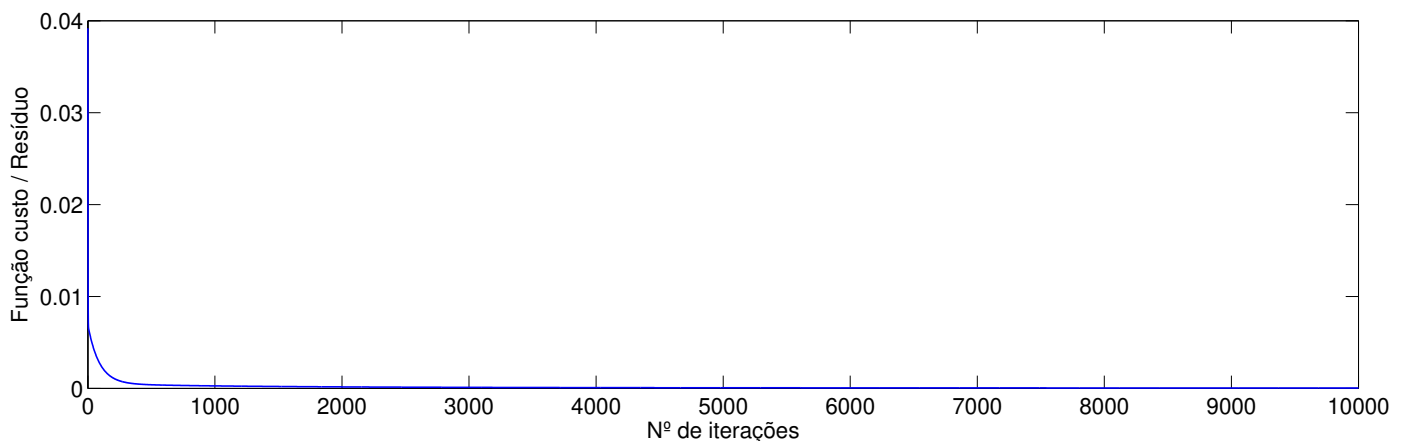
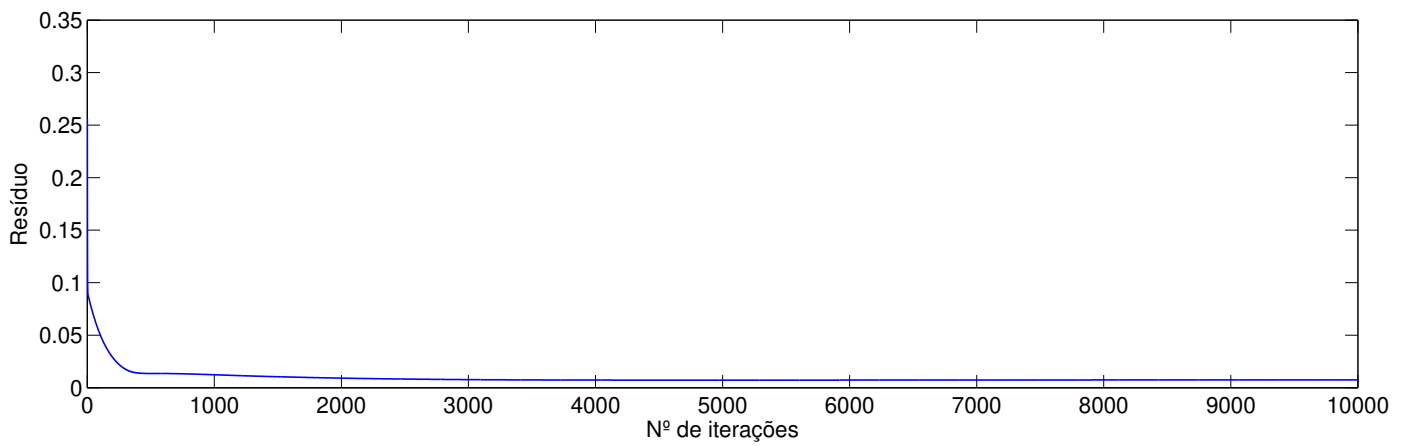


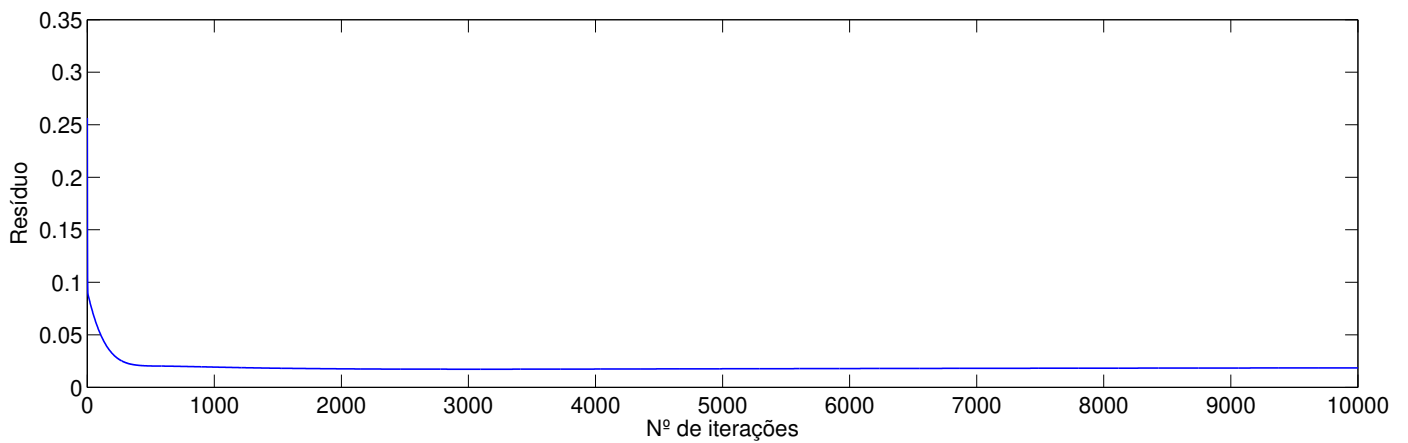
Figura 26 – Função custo conforme cada iteração no treinamento da técnica de regressão linear, através do método de Landweber, no conjunto do padrão de escoamento estratificado.

Seguindo o processo, aplicamos os vetores obtidos no conjunto de treinamento no conjunto de CV, porém agora os dados deste conjunto apresentam ruído. O procedimento adotado foi o seguinte: criamos 10 cópias idênticas da matriz de CV, e aplicamos a cada uma das linhas dessas matrizes o mesmo nível de ruído δ conforme (7.16). Após isso, para cada uma das 10 matrizes, calculamos uma sequência a_k , conforme (7.6), e fazemos a média de cada uma das entradas dessas sequências. Os resultados para os três níveis de ruído estão apresentados na Figura 27. A iteração onde o mínimo é atingido será o \hat{k} .

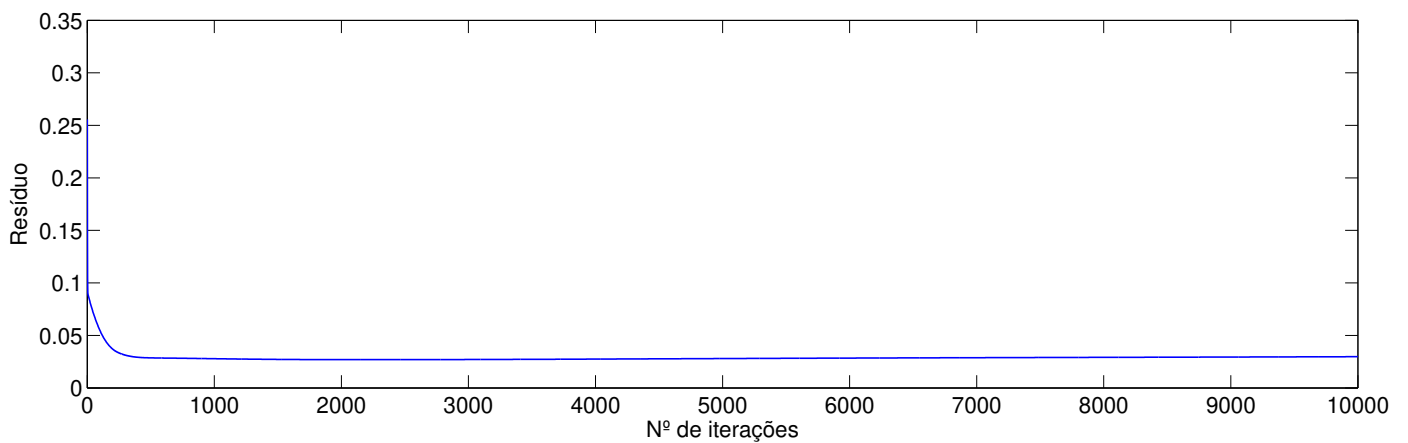
Mesmo que graficamente a mudança não seja tão nítida, é possível perceber que o mínimo atingido pela média das sequências a_k aumenta com o aumento do nível de ruído. Além disso, esse mínimo é atingido num índice menor, conforme o nível de ruído aumenta.



(a) 10% de ruído



(b) 30% de ruído



(c) 50% de ruído

Figura 27 – *Cross Validation* das regressões lineares para o padrão de escoamento estratificado, com diferentes níveis de ruído.

Comparando os resultados dos três níveis de ruído aplicados aos dados do padrão de escoamento estratificado, com os dados exatos no conjunto teste, obtemos a Figura 28.

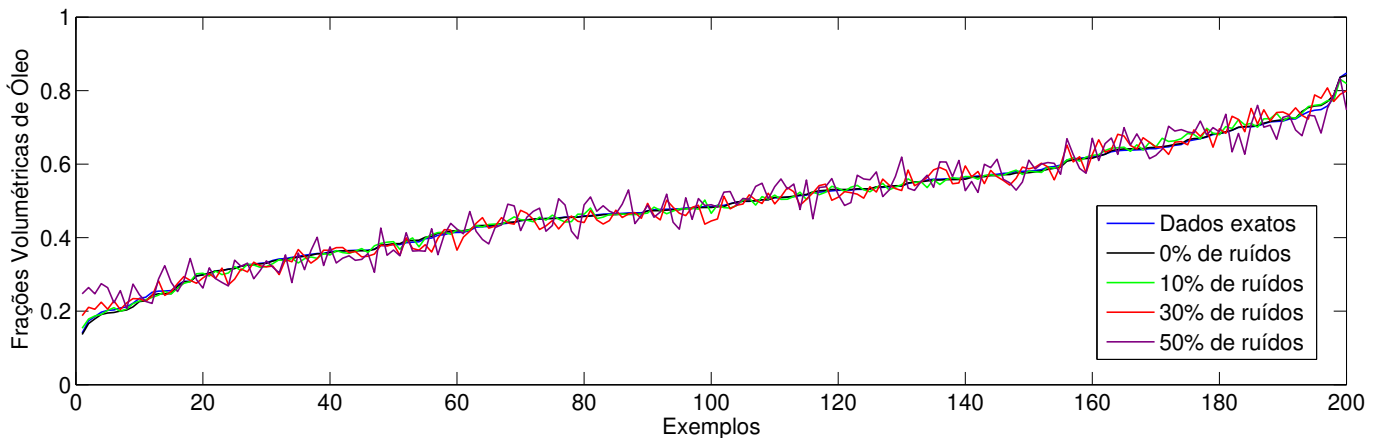


Figura 28 – Conjunto teste: Frações volumétricas obtidas com diferentes níveis de ruído, para o padrão de escoamento estratificado.

Para analisarmos de forma mais precisa, podemos observar na Tabela 2, a média dos erros no conjunto teste, isto é, a média dos módulos das diferenças entre as frações volumétricas calculadas e as frações exatas de cada nível de ruído. Além disso, a mesma tabela mostra a iteração que apresentou melhor desempenho no CV.

Padrão Estratificado	Ruído	Média dos erros	Melhor Iteração
	0%	0,0028	10000
	10%	0,0070	5285
	30%	0,0163	2968
	50%	0,0290	1706

Tabela 2 – Comparação da média de erros no cálculo das frações volumétricas e melhor iteração para cada nível de ruído, para o padrão de escoamento estratificado.

Podemos notar a partir da Tabela 2, que conforme o nível de ruído aumenta, a média do erro no cálculo das frações volumétricas também aumenta. Também pode-se notar que o índice da melhor iteração aumenta conforme o nível de ruído diminui. Isso porque, se recordarmos da Seção 3.2, espera-se que o erro tenda a 0 e $\hat{k} \rightarrow \infty$, quando $\delta \rightarrow 0$.

Para os demais padrões, podemos observar nas Tabelas 3 e 4, comportamentos semelhantes aos vistos no padrão estratificado.

Padrão Mistura	Ruído	Média dos erros	Melhor iteração
	0%	0,0120	10000
	10%	0,0126	10000
	30%	0,0188	10000
	50%	0,0306	4111

Tabela 3 – Comparação da média de erros no cálculo das frações volumétricas e melhor iteração para cada nível de ruído, para o padrão de escoamento mistura.

Padrão Anular	Ruído	Média dos erros	Melhor Iteração
	0%	0,0093	10000
	10%	0,0122	10000
	30%	0,0242	2780
	50%	0,0356	1471

Tabela 4 – Comparação da média de erros no cálculo das frações volumétricas e melhor iteração para cada nível de ruído, para o padrão de escoamento anular.

7.3.2 Determinação do padrão de escoamento

Nesta subseção, de maneira similar a anterior, são utilizados os vetores obtidos no treinamento da Seção 8.2.

O procedimento executado no conjunto de *Cross Validation*, também é o mesmo executado na Seção 8.2, com a diferença do procedimento de inserção artificial de ruído, que é idêntico ao da subseção anterior.

Apresentamos na Figura 29 a evolução dos resíduos no conjunto de *Cross Validation* do padrão de escoamento anular, para diferentes níveis de ruído.

Podemos observar nos resíduos do conjunto CV do padrão anular, que conforme o nível de ruído aumenta, o resíduo em todas as iterações também aumenta. Em particular, podemos observar isso mais nitidamente na iteração que apresenta menor resíduo para os três níveis de ruído, que é a iteração de índice $\hat{k} = 10000$, em que conforme o aumento do ruído, o resíduo da iteração $\hat{k} = 10000$ se aproxima de 0,05.

Dando sequência, selecionados os vetores com menor resíduo, para cada padrão de escoamento, executamos o processo de classificação para os dados com ruído do conjunto teste e, da mesma forma que na Seção 8.2, construímos as Tabelas 5, 6 e 7.

10% de ruído			
Calculados \ Exatos	Padrão Mistura	Padrão Estratificado	Padrão Anular
Padrão Mistura	200	0	0
Padrão Estratificado	0	200	0
Padrão Anular	0	0	200

Tabela 5 – Matriz confusão para a taxa de 10% de ruído.

30% de ruído			
Calculados \ Exatos	Padrão Mistura	Padrão Estratificado	Padrão Anular
Padrão Mistura	200	0	0
Padrão Estratificado	1	199	0
Padrão Anular	0	0	200

Tabela 6 – Matriz confusão para a taxa de 30% de ruído.

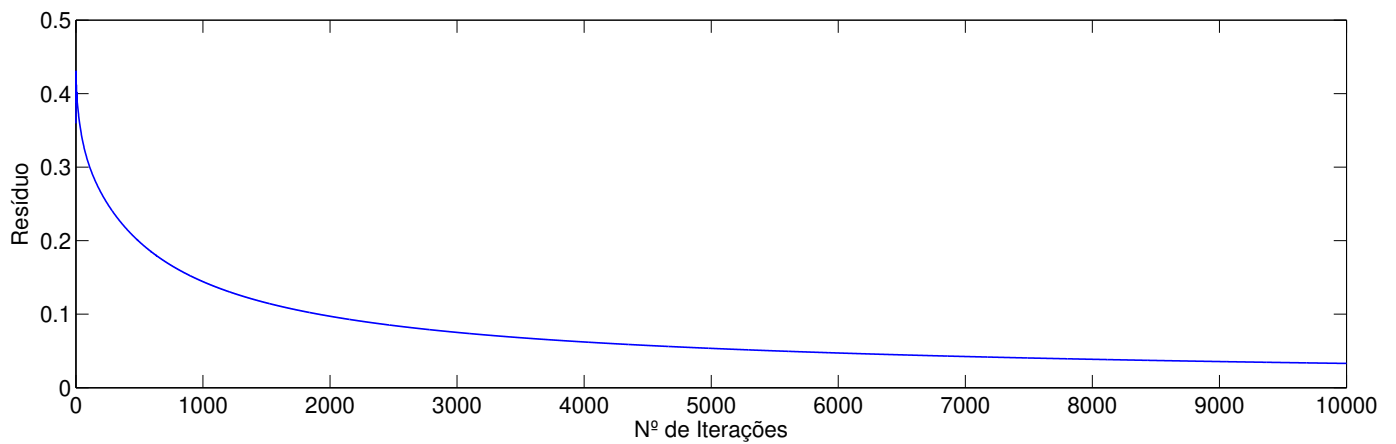
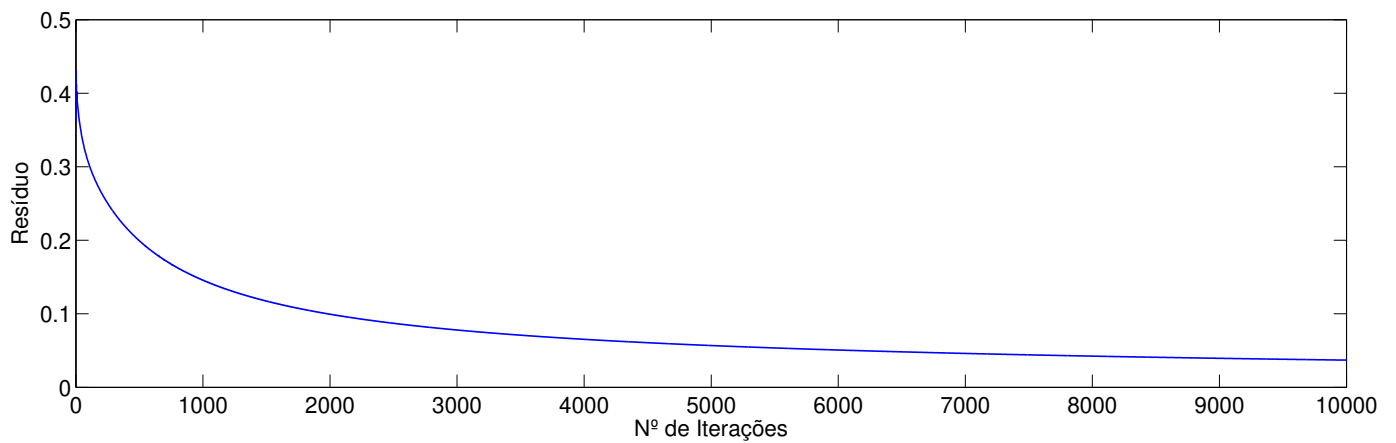
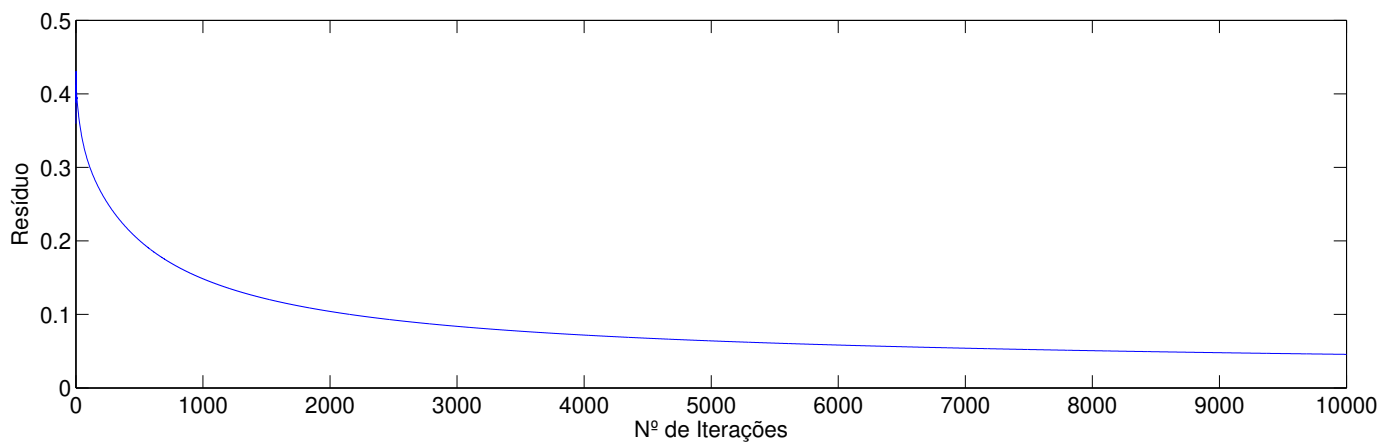
(a) *Cross Validation* com 10% de ruído(b) *Cross Validation* com 30% de ruído(c) *Cross Validation* com 50% de ruído

Figura 29 – *Cross Validation* das regressões logísticas para o padrão de escoamento anular, com diferentes níveis de ruído.

Podemos notar que a medida que o nível de ruído aumenta, fica mais difícil para o algoritmo classificar corretamente entre os padrões de escoamento, mas mesmo assim o algoritmo apresentou um bom desempenho, pois com a taxa de 50% de nível de ruído ele

50% de ruído			
Calculados \ Exatos	Padrão Mistura	Padrão Estratificado	Padrão Anular
Padrão Mistura	196	0	4
Padrão Estratificado	0	200	0
Padrão Anular	0	0	200

Tabela 7 – Matriz confusão para a taxa de 50% de ruído.

erra a classificação de 4 exemplos dentre os 600 do conjunto teste.

8 Considerações Finais

Neste trabalho de conclusão de curso foram realizadas as implementações numéricas dos algoritmos de *Machine Learning* de Regressão Linear e Regressão Logística para a determinação de fração volumétrica e classificação de padrões de escoamento, respectivamente, através de dados simulados do problema inverso de tomografia por impedância elétrica aplicada a tubos de extração de petróleo. Também fizemos a mesma implementação para os dados simulados com ruído inserido artificialmente. Os treinamentos dos algoritmos de Regressão Linear sem ruído foram realizados através dos métodos de Landweber e Tikhonov. Já os treinamentos das Regressões Logísticas, tanto com quanto sem ruído, e o treinamento da Regressões Linear com ruído, foram feitos através do método de Landweber.

Os resultados obtidos foram satisfatórios para as duas técnicas de *Machine Learning* utilizadas, tanto no caso sem ruído quanto no caso com os ruído. Isso porque, no caso das determinações de frações volumétricas, foram obtidos erros médios relativamente pequenos tanto para os dados simulados sem ruído, quanto para os dados simulados com ruído. Já na determinação dos padrões de escoamento, a taxa de acerto se manteve alta (99,33%), mesmo com 50% de ruído.

Tendo em vista os resultados obtidos, podemos dar continuidade aos estudos testando o desempenho das técnicas utilizadas em dados reais. Além disso, podem ser exploradas técnicas de *Machine Learning* mais avançadas, como redes neurais, para a determinação simultânea do padrão de escoamento e da fração volumétrica, inclusive para escoamentos com mais de 2 fases.

Referências

- AGUILAR, J. C. Z. *Estudos numéricos para o problema da tomografia por impedância elétrica*. Tese (Doutorado) — Universidade de São Paulo, 2009.
- BERRAR, D. Cross-validation. In: _____. [S.l.: s.n.], 2018. ISBN 9780128096338.
- BORCEA, L. Electrical impedance tomography. *Inverse Problems*, IOP Publishing, v. 18, n. 6, p. R99–R136, oct 2002. Disponível em: <https://doi.org/10.1088/0266-5611/18/6/201/>.
- BREZIS, H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. [S.l.]: Springer New York, 2010. (Universitext). ISBN 9780387709130.
- DEISENROTH, M. P.; FAISAL, A. A.; ONG, C. S. *Mathematics for Machine Learning*. Cambridge University Press, 2020. Disponível em: <https://mml-book.com/>.
- GÉRON, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 1st. ed. [S.l.]: O'Reilly Media, Inc., 2017. ISBN 1491962291.
- KIRSCH, A. *An Introduction to the Mathematical Theory of Inverse Problems*. [S.l.]: Springer, 2011.
- MARGOTTI, F. J. Métodos de tikhonov. Disponível em: <https://fabiomargotti.paginas.ufsc.br/pesquisa/problemas-inversos/teoria-da-regularizacao/metodos-de-tikhonov/>.
- MARGOTTI, F. J. *On Inexact Newton Methods for Inverse Problems in Banach Spaces*. Tese (Doutorado) — Karlsruher Institut für Technologie, 2015.
- MITCHELL, T. *Machine Learning*. McGraw-Hill, 1997. (McGraw-Hill International Editions). ISBN 9780071154673. Disponível em: <https://books.google.com.br/books?id=EoYBngEACAAJ>.
- MOTA, F. R. M. *Tomometria Capacitiva Aplicada à Medição de Fração de Água em Escoamentos Bifásicos*. Tese (Doutorado) — Universidade Federal de Santa Catarina, 2015.
- MOTA, F. R. M. da. *Contribuição para o Desenvolvimento de Um Sistema de Medição Multifásica com Estágio de Separação Parcial*. 2011.
- NG, A. *Curso de Machine Learning*. [S.l.]. Disponível em: <https://www.coursera.org/learn/machine-learning/>.
- SOMERSALO, E.; CHENEY, M.; ISAACSON, D. Existence and uniqueness for electrode models for electric current computed tomography. *SIAM Journal on Applied Mathematics*, Society for Industrial and Applied Mathematics, v. 52, n. 4, p. 1023–1040, 1992. ISSN 00361399. Disponível em: <http://www.jstor.org/stable/2102189/>.
- TARANTOLA, A. Inverse problems. Disponível em: <http://mathworld.wolfram.com/InverseProblem.html>.

APÊNDICE A – Espaços Normados e Espaços de Hilbert

Neste Apêndice foram coletadas algumas definições e resultados importantes da Análise Funcional, baseados no livro de Kirsch (2011).

Definição 9. (Produto Interno) *Seja X um espaço vetorial sobre o corpo $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$. O produto interno é uma função dada por:*

$$\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{K}$$

com as seguintes propriedades:

- $\langle x + y, z \rangle$,
- $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ para todo $x, y \in X$ e $\alpha \in \mathbb{K}$,
- $\langle x, y \rangle = \overline{\langle y, x \rangle}$ para todo $x, y \in X$,
- $\langle x, x \rangle \in \mathbb{R}$ e $\langle x, x \rangle \geq 0$, para todo $x \in X$,
- $\langle x, x \rangle > 0$ se $x \neq 0$.

Um espaço vetorial X sobre \mathbb{K} com produto interno $\langle \cdot, \cdot \rangle$ é chamado de *Espaço Pré-Hilbert* sobre \mathbb{K} .

As seguintes propriedades podem ser derivadas da definição:

- $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ para todo $x, y, z \in X$,
- $\langle x, \alpha y \rangle = \bar{\alpha} \langle x, y \rangle$ para todo $x, y \in X$ e $\alpha \in \mathbb{K}$.

Definição 10. (Norma) *Seja X um espaço vetorial sobre o corpo $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$. Uma norma em X é uma função:*

$$\|\cdot\| : X \rightarrow \mathbb{R}$$

com as seguintes propriedades:

- $\|x\| > 0$ para todo $x \in X$ com $x \neq 0$,
- $\|\alpha x\| = |\alpha| \|x\|$ para todo $x \in X$ e $\alpha \in \mathbb{K}$,
- $\|x + y\| \leq \|x\| + \|y\|$ para todo $x, y \in X$.

Um espaço vetorial X sobre o corpo \mathbb{K} com norma $\|\cdot\|$ é chamado de *Espaço Normado* sobre \mathbb{K} .

A terceira propriedade da última definição é chamada de desigualdade triangular.

Teorema 6. *Seja X um espaço pré-Hilbert. A função $\|\cdot\| : X \rightarrow \mathbb{R}$ definida por:*

$$\|x\| := \sqrt{\langle x, x \rangle}, \quad x \in X$$

é uma norma, isto é, possui as propriedades da 10. Além disso,

- $|\langle x, y \rangle| \leq \|x\| \|y\|$ para todo $x, y \in X$ (desigualdade de Cauchy-Schwarz),
- $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2 \pm 2\operatorname{Re}\langle x, y \rangle$ para todo $x, y \in X$,
- $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$ para todo $x, y \in X$.

Todo espaço normado carrega consigo uma topologia induzida pela norma, isto é, podemos definir conjuntos abertos, fechados e compactos; seqüências convergentes; funções contínuas; etc. Introduzimos a definição de bolas de raio r e centro $x \in X$ por:

$$K(x, r) := \{y \in X : \|y - x\| < r\}, \quad K[x, r] := \{y \in X : \|y - x\| \leq r\}.$$

Definição 11. *Seja X um espaço normado sobre o corpo $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .*

1. *Um subconjunto $M \subset X$ é dito ser limitado se existe $r > 0$ com $M \subset K(x, r)$. O conjunto $M \subset X$ é dito ser aberto se para todo $x \in M$ existe $\epsilon > 0$ tal que $K(x, \epsilon) \subset M$. O conjunto $M \subset X$ é dito ser fechado se seu complemento $X \setminus M$ é aberto.*
2. *Uma seqüência $(x_k)_k \subset X$ é dita ser limitada se existe $c > 0$ tal que $\|x_k\| \leq c$ para todo k . A seqüência $(x_k)_k \subset X$ é dita ser convergente se existe $x \in X$ tal que $\|x - x_k\|$ converge a zero em \mathbb{R} . Denotamos o limite por $x = \lim_{k \rightarrow \infty} x_k$, ou escrevemos $x_k \rightarrow x$ conforme $k \rightarrow \infty$. A seqüência $(x_k)_k \subset X$ é chamada de seqüência de Cauchy se para todo $\epsilon > 0$ existe $N \in \mathbb{N}$ tal que $\|x_m - x_k\| < \epsilon$ para todo $m, k \geq N$.*
3. *Seja $(x_k)_k \subset X$ seja uma seqüência. $x \in X$ é dito ser um ponto de acumulação se existe uma subseqüência $(a_{k_n})_n$ que converge para x .*
4. *Um conjunto $M \subset X$ é dito ser compacto se para toda seqüência em M existe um ponto de acumulação em M .*

É possível provar que se o conjunto M é fechado se, e somente se, o limite de toda seqüência convergente $(x_k)_k \subset M$ também pertence a M . Os conjuntos:

$$\text{int}(M) := \{x \in M : \text{existe } \epsilon > 0 \text{ com } K(x, \epsilon) \subset M\}$$

e

$$\text{cl}(M) := \left\{ x \in X : \text{existe } (x_k)_k \subset M \text{ com } x = \lim_{k \rightarrow \infty} x_k \right\}$$

são chamados de interior e fecho, respectivamente, de M .

Teorema 7. *Sejam X um espaço normado sobre o corpo \mathbb{K} e $M \subset X$ um subconjunto.*

1. *M é fechado se, e somente se, $M = \text{cl}(M)$, e M é aberto se, e somente se, $M = \text{int}(M)$.*
2. *Se $M \neq X$ é um subespaço, então $\text{int}(M) = \emptyset$, e $\text{cl}(M)$ também é um subespaço.*
3. *Em espaços de dimensão finita todo subespaço é fechado.*
4. *Todo conjunto compacto é fechado e limitado. Em espaços de dimensão finita a recíproca também vale (Teorema de Bolzano-Weierstrass): em um espaço normado de dimensão finita, todo conjunto fechado e limitado é um conjunto compacto.*

Uma propriedade crucial do conjunto de números reais é a de ser um espaço completo. Esta é uma afirmação que assumimos por sua necessidade em muitos resultados de análise funcional.

Definição 12. (Espaço de Hilbert e Espaço de Banach) *Um espaço normado X sobre o corpo \mathbb{K} é dito ser completo ou um espaço de Banach se toda sequência de Cauchy converge em X . Um espaço pré-Hilbert completo é chamado de espaço de Hilbert.*

Definição 13. (Complemento Ortogonal) *Seja X um espaço pré-Hilbert (sobre $\mathbb{K} = \mathbb{R}$ ou \mathbb{C}).*

1. *Dois elementos x e y são ditos ortogonais se $\langle x, y \rangle = 0$.*
2. *Seja $M \subset X$ um subconjunto. O conjunto:*

$$M^\perp := \{x \in X : \langle x, y \rangle = 0 \text{ para todo } y \in M\}$$

é chamado de complemento ortogonal de M .

M^\perp é sempre um subespaço fechado e $M \subset (M^\perp)^\perp$. Além disso, $A \subset B$ implica que $B^\perp \subset A^\perp$.

O seguinte teorema é um resultado de fundamental importância na teoria de espaços de Hilbert e se baseia principalmente na propriedade do espaço ser completo.

Teorema 8. Teorema de Projeção *Sejam X um espaço de Hilbert e $V \subset X$ um subespaço fechado. Então $V = (V^\perp)^\perp$. Todo $x \in X$ possui uma única decomposição da forma $x = v + w$, em que $v \in V$ e $w \in V^\perp$. O operador $P : X \rightarrow V$, $x \mapsto v$, é chamado de operador de projeção ortogonal em V e tem as seguintes propriedades:*

1. $Pv = v$ para todo $v \in V$, i.e., $P^2 = P$;
2. $\|x - Px\| \leq \|x - v'\|$ para todo $v' \in V$.

As propriedades anteriores implicam que $Px \in V$ é a melhor aproximação de $x \in X$ no subespaço fechado V .

APÊNDICE B – Operadores lineares e limitados e operadores compactos

Para este Apêndice, vamos considerar X e Y espaços normados e $A : X \rightarrow Y$ um operador linear.

Definição 14. (*Operador Limitado*) O operador linear A é dito ser limitado se existe $c > 0$ tal que:

$$\|Ax\| \leq c\|x\| \text{ para todo } x \in X.$$

A menor dessas constantes é chamada de norma de A , isto é:

$$\|A\| := \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Teorema 9. *As seguintes afirmações são equivalentes:*

1. A é limitado.
2. A é contínuo em $x = 0$, isto é, $x_j \rightarrow 0$ implica que $Ax_j \rightarrow 0$.
3. A é contínuo para todo $x \in X$.

Para o caso em que $Y = \mathbb{K}$, denotamos por $X' := \mathcal{L}(X, \mathbb{K})$ o espaço dual de X . Analogamente, $(X')'$ é chamado de espaço bidual de X . A injeção canônica $J : X \rightarrow (X')'$, definida por:

$$(Jx)l := l(x), \quad x \in X, l \in X',$$

é linear, limitada, injetiva e satisfaz $\|Jx\| = \|x\|$ para todo $x \in X$.

Definição 15. (*Espaço Reflexivo*) O espaço normado X é dito ser um espaço reflexivo se a injeção canônica é sobrejetiva, isto é, um isomorfismo que preserva a norma indo de X no bidual $(X')'$.

O seguinte resultado fornece uma importante caracterização de X' em espaços de Hilbert.

Teorema 10. (*Teorema Riesz-Fischer*) Seja X um espaço de Hilbert. Para todo $x \in X$, o funcional $f_x(y) := \langle y, x \rangle$, $y \in X$, define uma função limitada de X para \mathbb{K} , isto é, $f_x \in X'$. Além disso, para todo $f \in X'$ existe um único $x \in X$ com $f(y) = \langle y, x \rangle$ para todo $y \in X$ e

$$\|f\| := \sup_{y \neq 0} \frac{|f(y)|}{\|y\|} = \|x\|.$$

O último teorema implica que todo espaço de Hilbert é reflexivo. Também implica na existência de um único operador adjunto para todo operador linear limitado $A : X \rightarrow Y$.

Teorema 11. (Operador Adjunto) *Seja $A : X \rightarrow Y$ um operador linear e limitado entre espaços de Hilbert. Então, existe um único operador linear e limitado $A^* : Y \rightarrow X$ com a propriedade:*

$$\langle Ax, y \rangle = \langle x, A^*y \rangle \text{ para todo } x \in X, y \in Y.$$

O operador $A^* : Y \rightarrow X$ é chamado de operador adjunto de A . Se $X = Y$, o operador A é chamado de operador auto-adjunto se $A^* = A$.

Definição 16. (Operador Compacto) *O operador $K : X \rightarrow Y$ é dito ser compacto se para todo subconjunto limitado S , $K(S)$ é um conjunto relativamente compacto.*

Um conjunto $M \subset Y$ é dito ser relativamente compacto se para toda sequência limitada $(y_j) \subset M$ tem um ponto de acumulação em $cl(M)$, isto é, se o fecho $cl(M)$ é compacto.