Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 10, n. 1, 2023.

Trabalho apresentado no XLII CNMAC, Universidade Federal de Mato Grosso do Sul - Bonito - MS, 2023

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Experimentos de aprendizado por reforço para micro nadadores articulados virtuais

Luciano D. A. de Souza[1], Gustavo C. Buscaglia[2], Roberto F. Ausas[3]
ICMC/USP, São Carlos, SP
Stevens Paz[4]
Departamento de Matemáticas/Universidad del Valle, Calle, Cali, Colombia

Resumo. Nos últimos anos, os micros nadadores (biológicos ou sintéticos) têm atraído a atenção de muitos pesquisadores em todo o mundo devido às suas aplicações na medicina e na indústria. O desenvolvimento de robôs nadadores em micro ou nanoescala traz os benefícios de acessar locais pequenos e interagir com os elementos neste ambiente particular. Para poder nadar em micro escala, é necessário aprender estratégias de propulsão eficientes. Neste trabalho, discutiremos um modelo de elementos finitos de micro nadadores que serão treinados por um algoritmo de aprendizado por reforço para executar algumas tarefas simples. O modelo de elementos finitos resolve o problema de interação do fluido com a estrutura sólida e o algoritmo de aprendizado por reforço é o Q-Learning. Os resultados irão ilustrar estatisticamente o comportamento do micro nadador em cada uma das tarefas aprendidas.

Palavras-chave. Micro nadador, Elementos Finitos, Navier-Stokes, Aprendizado por Reforço, Q-Learning

1 Introdução

Na natureza, os microorganismos podem tirar proveito de estímulos ambientais para influenciar seus padrões de mobilidade, com o intuito de atingir algum objetivo biologicamente relevante [1]. A informação sobre o estado do ambiente é sentida, processada e codificada no microorganismo a fim de que possa fornecer ações ou propriedades futuras apropriadas para ele [1]. Com essa inspiração da natureza, os robôs em micro e nanoescalas passaram a ser projetados para executar tarefas especializadas em ambientes complexos [7].

Para a criação destes modelos de micro nadadores os métodos de elementos finitos possuem algumas vantagens. Esses métodos são prontamente estendidos para reologia não newtoniana e números de Reynolds diferentes de zero, enquanto os elementos de contorno dependem da linearidade do problema [6]. Além disso, os elementos finitos fornecem uma representação esparsa do campo de velocidade para cálculos de advecção, enquanto os elementos de contorno precisam passar por uma etapa de pós-processamento bastante custosa [6]. Neste trabalho, para o tratamento da interação dinâmica entre o fluido e o micro nadador, resolveremos as equações de Navier-Stokes utilizando elementos finitos de forma preditiva - isto é, com movimentos arbitrários - e não obstante, permitindo condições gerais de contorno na interface do problema fluído-estrutura.

Recentemente, a abordagem de aprendizado de máquina tem sido amplamente utilizada para investigar as estratégias comportamentais de matéria ativa em fluidos, visto que para alcançar

[1]lucianodellier@usp.br
[2]gustavo.buscaglia@icmc.usp.br
[3]rfausas@icmc.usp.br
[4]stevens.paz@correounivalle.edu.co

metas pré-definidas, o micro nadador precisa ajustar o seu movimento captando informações do fluxo local [2]. Além disso, esses estudos demonstram o grande potencial do aprendizado por reforço, em busca de formas eficazes de natação para minúsculas partículas ativas [2].

Para a implementação do modelo fluido-estrutura, utilizaremos a plataforma de elementos finitos FEniCS. Para o processo de aprendizado por reforço, por sua vez, utilizaremos o algoritmo Q-Learning, que será implementado em Python. Finalmente, as atividades aprendidas serão rotacionar o corpo e nadar em uma direção específica [5]. Para a quantificação dos resultados obtidos, utilizaremos uma estrutura de micro nadador, cuja quantidade de corpos é variável - i.e. pode ser alterada -. Nesse sentido, o comportamento do micro nadador será apresentado consoante aos processos de aprendizado desenvolvidos. Ademais, dados estatísticos permitirão quantificar a qualidade e a eficiência do processo de aprendizado do micro nadador.

2 Formulação matemática do movimento da natação e seu controle

As equações que modelam a dinâmica de um fluido são dadas pelas equações de Navier-Stokes. Como estamos interessados em fluidos em regime com baixo Reynolds, onde há domínio das forças viscosas sobre as inerciais, podemos desconsiderar o termo inercial, obtendo assim um problema de Stokes, que com a condição de incompressibilidade, é dado por

$$-\nabla \cdot \sigma(\mathbf{u}) = 0, \quad \nabla \cdot \mathbf{u} = 0 \tag{1}$$

em que $\mathbf{u}$ é a velocidade e σ é o tensor de tensões de Cauchy, que para fluidos newtonianos, é dado por

$$\sigma = -pI + \mu(\nabla \mathbf{u} + (\nabla \mathbf{u})^T) \tag{2}$$

onde p é a pressão, μ é a viscosidade do fluido e I a matriz identidade.

Neste contexto, consideramos um modelo de micro nadador tal como demonstrado na figura 1 cuja forma corporal, a menos de um movimento rígido, esteja totalmente definida por um vetor de m variáveis internas (ângulos entre membros, comprimento de braços, etc.), que denotaremos como $\xi \in \mathbb{R}^m$. Na hipótese mais simples, essas variáveis não têm restrições, e podem assumir qualquer valor real.

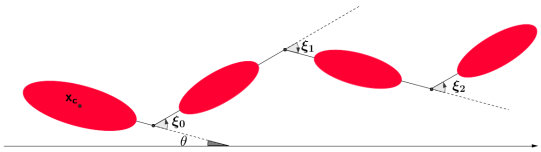

Figura 1: Estrutura do micro nadador composto por quatro corpos, no qual a cabeça (à esquerda) é maior que os demais membros. O ângulo entre a cabeça e o eixo horizontal é denotado por θ e os ângulos entre os membros são denominados por ξ. Fonte: [4]

A cada instante de tempo, t, o nadador terá a forma correspondente às variáveis $\xi(t)$, e estará posicionado num ponto $\mathbf{x}_c(t)$ no espaço (no referencial do laboratório) e orientado segundo um ângulo $\theta(t)$ com respeito a uma orientação de referência. Em particular, estamos considerando um caso bidimensional (2D), onde a orientação requer apenas um ângulo, o qual é definido com base no movimento de translação do centro de massa $\mathbf{c}$ do corpo.

DOI: 10.5540/03.2023.010.01.0052

Supomos ainda que o nadador tem completo controle das variáveis internas, portanto, pode especificar um histórico (contínuo) de movimentos $\xi : [0, T] \to \mathbb{R}^m$. Logo, podemos enunciar o primeiro problema.

Problema 1 (predição): Dados $\mathbf{x_c}(0)$ e $\theta(0)$ e um histórico de movimentos $\xi \in C^0(0, T, \mathbb{R}^m)$, determinar $\mathbf{x}_c(t)$ e $\theta(t)$ para $t \in (0, T]$.

Esse é um problema bem posto se não existem colisões entre as partes do corpo. É um problema conceitualmente simples, mas cuja formulação matemática requer algumas definições específicas. Primeiramente, seja $\mathbf{p} = (\mathbf{x}_c, \theta)$ o vetor de variáveis externas do nadador e seja $\mathbf{q} = (\mathbf{p}, \xi)$ o vetor de todas as variáveis que determinam a posição e orientação de cada parte material dele. Os pontos fundamentais do modelo matemático são:

- Para cada $\mathbf{q} = (\mathbf{x}_c, \theta, \xi)$ e cada $\dot{\mathbf{q}} = (\dot{\mathbf{x}}_c, \dot{\theta}, \dot{\xi})$ é induzido um campo vetorial $\mathbf{v}$ na superfície do corpo (que denotaremos por Γ). Note que no problema proposto $\dot{\xi}$ é uma função conhecida mas $\dot{\mathbf{x}}_c$ e $\dot{\theta}$ devem ser consideradas incógnitas;
- O campo de velocidade $\mathbf{u}$ do fluido deve coincidir com $\mathbf{v}$ em Γ. Essa é a condição de aderência (ou "stick") entre o fluido e o sólido;
- O campo de velocidade $\mathbf{u}$ do fluido, com o campo de pressão p e o campo de tensão σ, devem satisfazer as equações de Stokes (1)-(2);
- A força e o torque que o fluido exerce sobre o nadador devem ser iguais à força e ao torque externos ($\mathbf{f}^{\text{ext}}$ e $\mathbf{t}^{\text{ext}}$) aplicados sobre ele (usualmente ambos iguais a zero). Em termos do campo de tensão $\sigma(\mathbf{u})$, este se expressa como:

$$\int_\Gamma \sigma(\mathbf{u}) \cdot \mathbf{n} \, dS = \mathbf{f}^{\text{ext}}, \tag{3}$$

$$\int_\Gamma ((\mathbf{x} - \mathbf{x}_c) \times \sigma(\mathbf{u}) \cdot \mathbf{n} \, dS = \mathbf{t}^{\text{ext}}. \tag{4}$$

  De uma maneira bastante implícita, (3)-(4) são 3 equações (em 2D) que determinam as incógnitas $\dot{\mathbf{x}}_c$ e $\dot{\theta}$. Integrando no tempo essas últimas, temos a solução do Problema 1.

Note que, em cada instante de tempo, as equações (3)-(4) envolvem a solução de um problema de Stokes com condição de contorno $\mathbf{v}$ na superfície do corpo. A formulação variacional do problema pode ser consultada em [3].

Surge então o **problema inverso**, em particular na sua formulação de **controle ótimo**. Isto é, temos a pergunta sobre qual é a sequência $\xi(t)$ de movimentos que maximiza alguma medida de eficiência ou **função objetivo** $J[\xi]$ (uma função real cuja variável é uma função $\xi \in C^0(0, T, \mathbb{R}^m)$.

Problema 2 (controle): Seja $J : C^0(0, T, \mathbb{R}^m) \to \mathbb{R}$ uma função contínua. Determine $\xi^* : (0, T) \to \mathbb{R}^m$ que maximize J.

A obtenção da solução do Problema 2, ou melhor, uma aproximação dela, pode ser encarada como um problema de **aprendizado**. Para isto, o tempo é discretizado em intervalos regulares. Estes passos de tempo restringem as funções ξ admissíveis e assim facilitam o problema de otimização. Se $\mathbf{q}(t)$ é o estado do agente a tempo t, a restrição imposta é que em cada passo de tempo $(t, t + 1)$ o nadador (ou **agente**) apenas pode realizar **uma ação** a, dentre um conjunto A_t de ações possíveis.

Com essa discretização da variável temporal o Problema 2 se transforma em determinar a sequência de ações $a_0, a_1, \ldots$ (dentre as possíveis) que maximiza J. Note que, mesmo que as possíveis ações a cada instante sejam finitas e até poucas, as sequências possíveis crescem muito rapidamente com o número de passos de tempo ao longo dos quais está sendo resolvido o Problema 2. Por isto, são necessários algoritmos especiais para determinar o máximo de J, em particular

DOI: 10.5540/03.2023.010.01.0052

algoritmos de programação dinâmica que são conhecidos na Ciência de Dados como algoritmos de **aprendizado por reforço (AR)**.

3 Aprendizado por reforço

A aprendizagem por reforço é um treinamento de modelos de aprendizado de máquina, que mapeia as situações das ações a serem executadas, o agente do aprendizado não é informado sobre quais ações deve se tomar, mas deve descobrir quais ações rendem maior recompensa ao experimentá-las.

Na linguagem do AR, o processo de tomada de decisões é modelado como uma **política de controle** $\pi(a(t)|S(t), \psi(t))$, que expressa a probabilidade de escolher a ação $a(t)$ quando o estado do agente é $S(t)$. Incluímos uma nova variável de estado ψ para denotar quaisquer variáveis internas do próprio sistema nervoso do agente, ou, em termos computacionais, variáveis internas do algoritmo de decisão.

O agente, como resultado de cada ação, não só modifica seu estado de $S(t)$ a $S(t+1)$, como também recebe uma **recompensa** $r(t+1)$. A **recompensa total** é calculada como $R = \sum_{t=1}^{T} \gamma^t r(t)$, onde $\gamma \in (0, 1]$ é um fator de desconto que determina o grau de preferência de recompensas imediatas sobre recompensas futuras.

3.1 O algoritmo de Q-Learning

Q-Learning é um algoritmo de aprendizado por reforço que busca determinar a melhor ação a ser tomada, dado o seu estado atual. As variáveis aprendidas $\psi(t)$ são organizadas na forma de uma **matriz ação-valor** $Q(t)$. O número de linhas de $Q(t)$ é igual ao número de estados possíveis, e o número de colunas é igual ao número de ações possíveis.

Utilizando a versão assíncrona do algoritmo de Q-Learning, o micro nadador produz uma série de experiências exploratórias onde cada ação $a(t)$ tomada, no estado $S(t)$ - e no fim da ação acabou no estado $S(t + 1)$ - a matriz Q é atualizada como $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(r_{t+1} + \gamma \max_{a'} Q(S_{t+1}, a') - Q(S_t, A_t))$.

Essa particular atualização de Q corresponde a uma iteração de valor, na linguagem do AR, e sob hipóteses adequadas converge à solução da equação de otimalidade de Bellman [5]. O coeficiente de relaxação α é conhecido como taxa de aprendizado.

Note que o cálculo hidrodinâmico está presente nessa descrição de maneira sutil, já que ele aparece implicitamente no cálculo do novo estado $S(t + 1)$ e da recompensa $r(t + 1)$. Para mais detalhes do algoritmo consultar em [5].

4 Experimentos numéricos

O primeiro experimento com aprendizado por reforço, para a estrutura de micro nadador exemplificado na figura 1, é a rotação do seu corpo. A recompensa é definida como $r(t) = \theta(t+1) - \theta(t)$, onde $\theta(t)$ é o ângulo da cabeça antes de executar a ação e $\theta(t + 1)$ é o ângulo da cabeça após executar a ação. Devido às características do modelo do micro nadador, e a forma da solução do problema do movimento da natação, todos estados e ações possíveis são conhecidos. Assim, podemos atualizar cada estado e ação, da matriz de Q-Learning, em um único passo de aprendizado, a fim de encontrar a matriz exata para cumprir a nossa tarefa. O objetivo da figura 2, é mostrar como a quantidade de corpos influencia na quantidade de atualizações da matriz Q, enquanto a figura 3 demonstra a influência que γ tem sobre o aprendizado para cada tipo de micro nadador, e por fim a figura 5 ilustra a política aprendida para o micro nadador com 10 corpos.

DOI: 10.5540/03.2023.010.01.0052

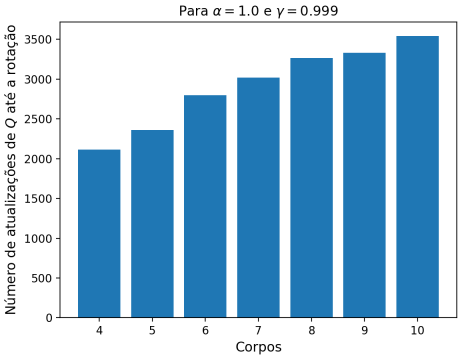

Figura 2: Número de atualizações da matriz Q por completo.

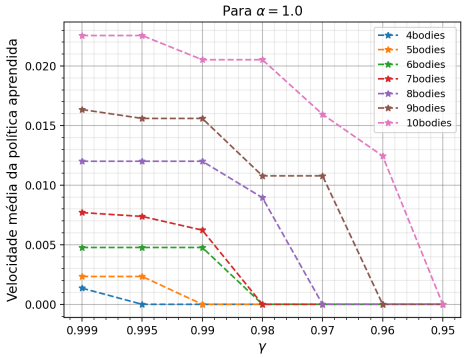

Figura 3: Velocidade média de rotação aprendida, para cada γ estudado.

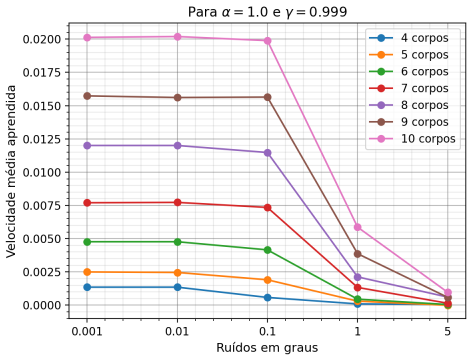

Figura 4: Comportamento do aprendizado de rotação com diversos tipos de ruídos no fluido.

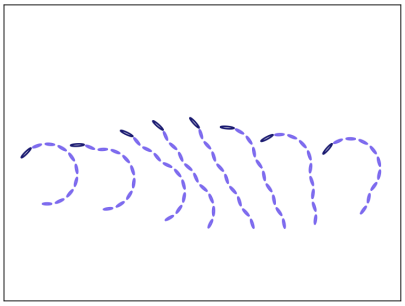

Figura 5: Política aprendida para a rotação do micro nadador de 10 corpos.

A figura 2 mostra que, quanto mais corpos a estrutura do micro nadador tiver, mais atualizações na matriz Q por completo serão necessárias para atingir o objetivo de rotacionar. A figura 3, no entanto, ilustra que quanto mais corpos os micro nadadores tiverem, menor a necessidade de um γ alto para aprender uma política de rotação.

Como estudo final sobre a rotação, a figura 4 ilustra uma análise na qual são geradas 100 trajetórias aleatórias de 5 milhões de experiências com diversos tipos de ruídos no fluido. Eles são produzidos utilizando a equação $\theta_{t+1} = \theta_t + \Delta(S_t, A_t) + \eta$, onde η tem distribuição normal com média 0 e desvio padrão igual aos valores representados pelo eixo das abscissas (eixo x) da imagem 4. O algoritmo de Q-Learning é aplicado sobre cada um dos caminhos explorados. A velocidade no eixo y representa a média das 100 velocidades aprendidas para cada ruído. E podemos concluir que, os diversos tipos de micro nadadores conseguem aprender a rotacionar com ruídos no fluido menores que 1 grau.

O segundo experimento numérico é nadar em uma direção específica. Para essa análise geramos experiências exploratórias, e delas estudamos todos os ciclos de ações em que um determinado estado do micro nadador se repete, e analisamos a distância percorrida - no referencial (x, y) - para cada um dos ciclos de ações. E será selecionado a política que obtiver a maior norma da sua distância, pois esta irá produzir o maior deslocamento linear. Portanto, para este caso, o algoritmo

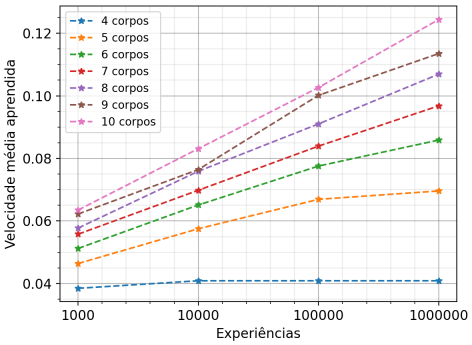

Figura 6: Velocidade média aprendida com relação a 100 variações das experiências analisadas.

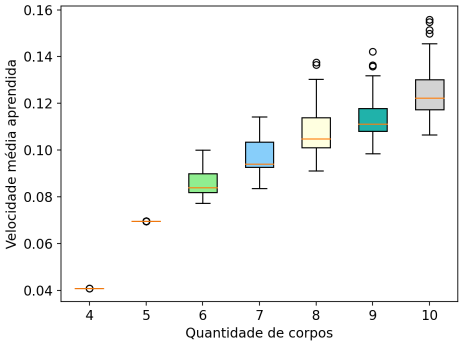

Figura 7: Análise estatística da velocidade média de 100 variações de 1 milhão de experiências.

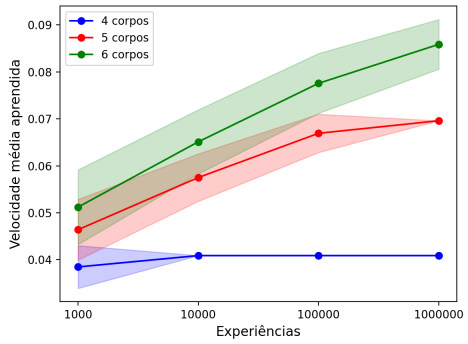

Figura 8: Média das velocidades médias aprendidas com relação à dispersão dos dados.

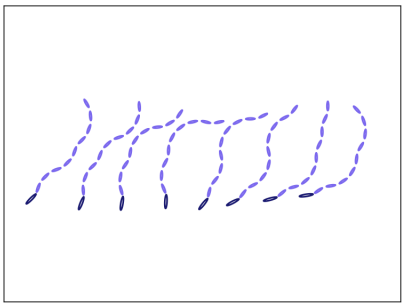

Figura 9: Política aprendida do nado em uma única direção do micro nadador de 10 corpos.

de Q-Learning não é utilizado.

A figura 6 tem como proposito ilustrar o comportamento deste aprendizado com relação à quantidade de experiências exploratórias, e as imagens 7 e 8 trazem a informação estatística de que micro nadadores com pouca quantidade de corpos, aprendem a nadar em uma direção com poucas experiências exploratórias, e a figura 9 ilustra a mais eficiente política aprendida - com 100 variações de 1 milhão de experiências exploratórias - para o micro nadador com 10 corpos.

A imagem 6 representa um estudo no qual são geradas 100 variações de trajetórias aleatórias com 1 mil, 10 mil, 100 mil e 1 milhão de experiências. Para cada um dos conjuntos de experiências, é calculada a média das velocidades médias que proporciona o maior deslocamento linear. Esse resultado mostra que o micro nadador de 4 corpos não necessita de muitas experiências para aprender a política ótima do nado em linha reta, enquanto o micro nadador de 10 corpos ainda apresenta um sequência de velocidades aprendidas de forma crescente.

A figura 7 mostra uma análise estatística das 100 variações de 1 milhão de experiências para um micro nadador com 4, 5, 6, 7, 8, 9 e 10 corpos. Desta forma, podemos concluir que com 8, 9 e 10 corpos, temos a presença de outliers que indicam que para essa quantidade de corpos, o processo de aprendizado ainda está longe de atingir a velocidade ótima desta tarefa.

DOI: 10.5540/03.2023.010.01.0052

Para complementar o estudo da imagem 7, a figura 8 mostra em linha contínua a média das velocidades aprendidas e a área sombreada acima e abaixo representa os valores da média mais o desvio padrão e a média menos o desvio padrão, respectivamente. Ela demonstra que o micro nadador de 4 e 5 corpos consegue aprender a velocidade ótima do aprendizado nas variações de experiências estudadas, enquanto o de 6 corpos ainda possui variações de velocidade.

5 Considerações Finais

Foram apresentados experimentos numéricos de aprendizado por reforço com micro nadadores articulados virtuais, cujo objetivo é ilustrar um modelo de elementos finitos para resolver o problema fluído-estrutura e aprender algumas tarefas simples.

Com isso, é possível concluir que, a presença de um número maior de corpos na estrutura do micro nadador leva a uma maior velocidade média de rotação, mas torna necessário um número maior de passo de aprendizado do algoritmo de Q-Learning. O aprendizado de rotação também será válido para fluidos com ruídos menores que 1 grau.

Noutra vertente, os resultados referentes ao nado em uma direção específica, mostram que o micro nadador consegue aprender a tarefa proposta sem a necessidade da utilização do Q-Learning, e que um micro nadador de 4 e 5 corpos consegue aprender a velocidade ideal para a realização desta tarefa.

Agradecimentos

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico e à Fundação de Amparo à Pesquisa do Estado de São Paulo(FAPESP).

Referências

[1] T. J. Pedley e J. O. Kessler. "Hydrodynamic Phenomena in Suspensions of Swimming Microorganisms". Em: **Annual Review of Fluid Mechanics** 24.1 (1992), pp. 313–358. DOI: 10.1146/annurev.fl.24.010192.001525.

[2] JingRan Qiu et al. "Swimming strategy of settling elongated micro-swimmers by reinforcement learning". Em: **Science China Physics, Mechanics & Astronomy** 63.8 (2020). DOI: 10.1007/s11433-019-1502-2.

[3] Stevens Sánchez e Gustavo Buscaglia. "Simulating squirmers with volumetric solvers". Em: **Journal of the Brazilian Society of Mechanical Science and Engineering** (2020).

[4] Paula Jaíne Alves da Silva et al. **Aprendizado de estratégias de propulsão de micro-nadadores a baixo número de Reynolds**. 2022. DOI: 10.5540/03.2022.009.01.0238.

[5] Richard S. Sutton e Andrew G. Barto. **Reinforcement Learning: An Introduction**. A Bradford Book, 2018. ISBN: 0262039249.

[6] C. Taylor e P. Hood. "A numerical solution of the Navier-Stokes equations using the finite element technique". Em: **Computers & Fluids** 1.1 (1973), pp. 73–100. DOI: https://doi.org/10.1016/0045-7930(73)90027-3.

[7] Joseph Wang e Wei Gao. "Nano/Microscale Motors: Biomedical Opportunities and Challenges". Em: **ACS nano** 6 (2012), pp. 5745–51. DOI: 10.1021/nn3028997.

