

YAPAY ÖĞRENME VE KISITLI OPTİMİZASYON

Öğrenme için Optimizasyon
Optimizasyon için Öğrenme

Genel bakış

- Öğrenme için optimizasyon : Yeni problem yapıları
 - *Kısıtlı optimizasyon*
 - Güncel : Konveks olmayan stokastik
 - *Sıfırıncı derece optimizasyon (DFO)*
- Optimizasyon için öğrenme : Veriye-dayalı matematiksel modelleme
 - *Öğrenilmiş (kısıt) fonksiyonlarla optimizasyon*

Klasik çerçeve : Konveks deterministik

The Interplay of Optimization and Machine Learning Research

Kristin P. Bennett

*Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12018, USA*

BENNEK@RPI.EDU

Emilio Parrado-Hernández

*Department of Signal Processing and Communications
University Carlos III de Madrid
Leganés (Madrid), 28911, Spain*

EMIPAR@TSC.UC3M.ES

- SVM – QP (dual) / nonsmooth convex (primal)
- Farklı ML uygulamalarında konveks optimizasyon problemleri
 - LP (e.g. configuration of a graphical model)
 - SDP (e.g. graph clustering for partitioning a data set)
 - SOCP (e.g. robust SVM training)

Güncel çerçeve : Konveks olmayan stokastik

Linearly Constrained Neural Networks

Johannes N. Hendriks^{a,*}, Carl Jidling^b, Adrian G. Wills^a, Thomas B. Schön^b

- Popüler bir fikir : ‘soft’ kısıtlar kullanmak
 - *Ancak optimizasyon alanında iyi bilinen bir gerçek şu ki ‘penalty’ fonksiyonlarını doğrudan minimize etmek genellikle kötü bir kısıtlı optimizasyon yöntemidir*

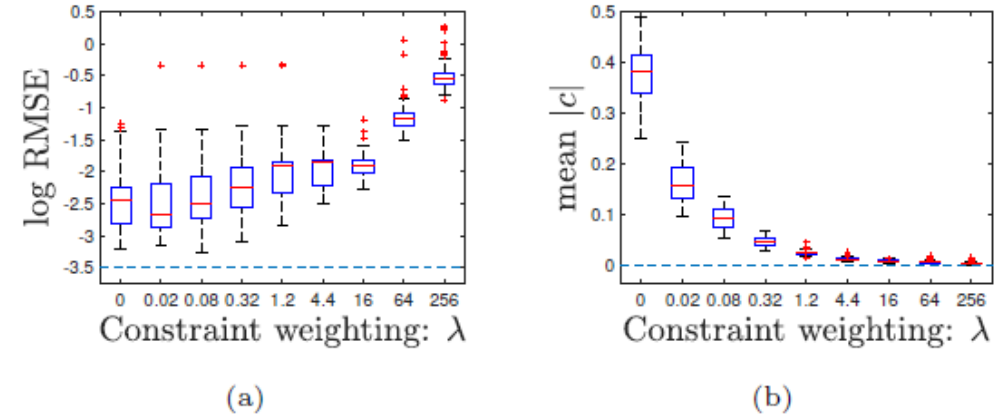


Figure 6: Performance of a neural network with cost function augmented by penalising the mean squared constraint violation. The study compares (a) the RMSE of the predicted field and (b) the mean absolute constraint violation of the predicted field for a range of weighting factors λ . For comparison, results from our proposed approach for the same number of measurements is indicated by the dashed line.

Yapısal kısıtlar

- Model parametrelerinin sağlaması ‘istenen’ yapısal özellikler
 - *Seyreklik, rank, pozitif-tanımlılık, vb*

NEW PENALIZED STOCHASTIC GRADIENT METHODS FOR LINEARLY CONSTRAINED STRONGLY CONVEX OPTIMIZATION

MENG LI*, PAUL GRIGAS*, AND ALPER ATAMTÜRK*

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & F(x) = \frac{1}{\ell} \sum_{i=1}^{\ell} f_i(x) + \psi(x) \\ \text{s.t.} \quad & a_i^T x \leq b_i, \quad i = 1, \dots, m. \end{aligned}$$

others. In machine learning, some examples include isotonic regression [7], convex regression [37, 25], and strong convex relaxations of sparse combinatorial problems such as signal estimation [5] and sparse regression [2, 3, 21]. We are particularly

Yapısal kısıtlar

- Öğrenme probleminin formülasyonu
- Multi-class Neyman-Pearson classification

$$\begin{aligned} \min_{\|x_k\| \leq \lambda, k \in [K]} \quad & \frac{1}{|\mathcal{D}_1|} \sum_{l>1} \sum_{\xi \in \mathcal{D}_1} h(x_1^T \xi - x_l^T \xi) \\ \text{s.t.} \quad & \frac{1}{|\mathcal{D}_k|} \sum_{l \neq k} \sum_{\xi \in \mathcal{D}_k} h(x_k^T \xi - x_l^T \xi) \leq \gamma_k, \quad k = 2, \dots, K, \end{aligned}$$

- K : sınıf sayısı
- \mathcal{D}_k : k .sınıf için eğitim verisi
- x_k : k .model parametreleri (sınıf tahmini için $\arg \max_{k \in [K]} x_k^T \xi$)
- Kayıp fonksiyonu : $h(z) = 1/(1 + e^z)$

Tanımsal kısıtlar (uzman bilgisi)

Bayesian Network Learning with Parameter Constraints

Radu Stefan Niculescu

*Computer Aided Diagnosis and Therapy Group
Siemens Medical Solutions
51 Valley Stream Parkway
Malvern, PA, 19355, USA*

STEFAN.NICULESCU@SIEMENS.COM

Tom M. Mitchell

*Center for Automated Learning and Discovery
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213, USA*

TOM.MITCHELL@CMU.EDU

R. Bharat Rao

*Computer Aided Diagnosis and Therapy Group
Siemens Medical Solutions
51 Valley Stream Parkway
Malvern, PA, 19355, USA*

BHARAT.RAO@SIEMENS.COM

- Uygulama alanındaki uzman, model parametrelerinin alabileceği değerler ve aralarındaki ilişkilere ilişkin bilgisini kısıt fonksiyonlarıyla ifade edebilir
- Genel kısıt fonksiyonlarından ziyade **özel kısıt yapılarını** (e.g. tek değişken ya da değişkenlerin toplamı için tanımlanmış kısıtlar) dikkate alarak kısıtlı problem çözmek daha etkin bir yaklaşım

Tanımsal kısıtlar (fiziksel denklemler)

- Modelin verdiği tahminler sistem hakkında bilinen fiziksel denklemleri tanım kümesinin her noktasında sağlayacak
- Daha az veri gerekecek
- Doğrusal eşitlik kısıtları gibi özel durumlarda problemin boyutu küçülecek

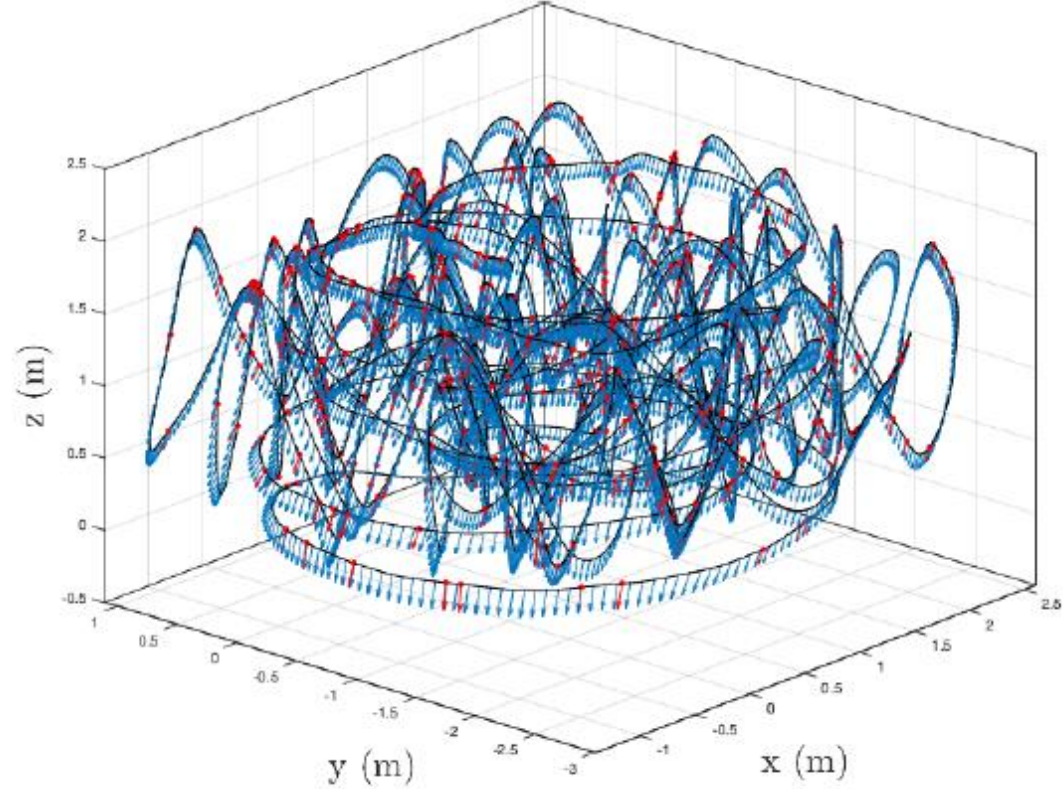


Figure 1: Magnetic field predictions (blue) using a constrained neural network trained on 500 observations (red) sampled from the trajectory indicated by the black curve. The magnetic field, \mathbf{B} is curl-free satisfying the constraint $\nabla \times \mathbf{B} = 0$, and the method proposed in this paper ensures that the predictions satisfy this constraint.

Tanımsal kısıtlar (fiziksel denklemler)

Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations (2019)

M. Raissi^a, P. Perdikaris^{b,*}, G.E. Karniadakis^a

■ PINN – Physics Informed Neural Networks

- *Kayıp fonksiyonunu sistem hakkında bilinen kısmi diferansiyel denklemleri sağlayacak şekilde minimize et*

■ Örnek : Akışkanlar dinamiği / 2D - Navier-Stokes denklemleri

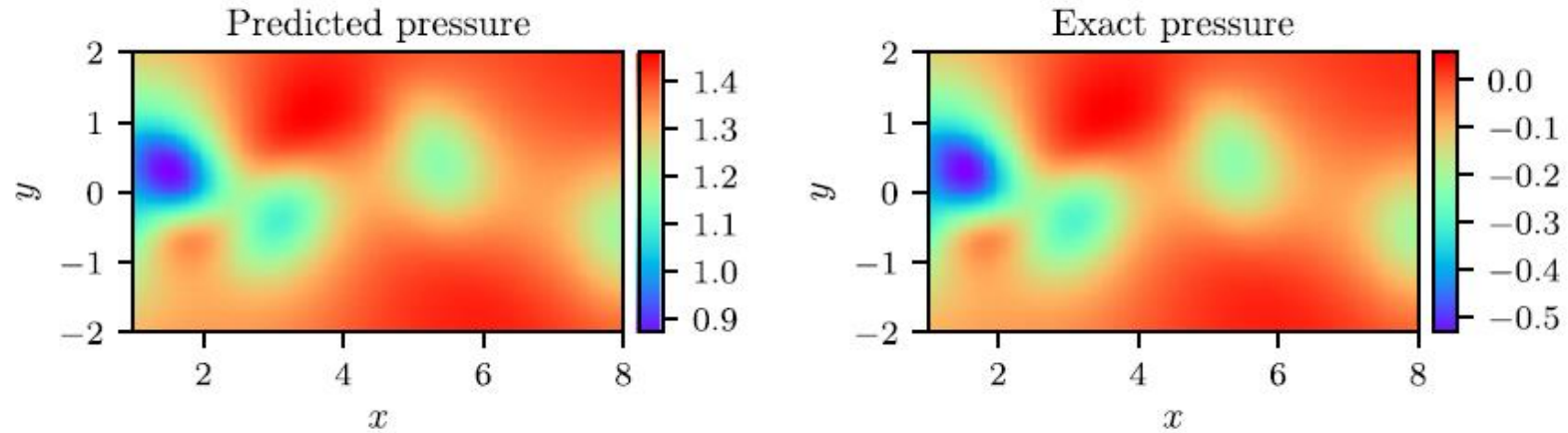
$$\begin{aligned} u_t + \lambda_1(uu_x + vu_y) &= -p_x + \lambda_2(u_{xx} + u_{yy}), & \longrightarrow f(t, x, y) \\ v_t + \lambda_1(uv_x + vv_y) &= -p_y + \lambda_2(v_{xx} + v_{yy}), & \longrightarrow g(t, x, y) \end{aligned}$$

- p : basınç, (u,v) : hız alanı (velocity field), λ : bilinmeyen parametreler
- Basınç ve parametreleri öğrenmek istiyoruz; elimizde hız alanı ölçümleri var $\{t^i, x^i, y^i, u^i, v^i\}_{i=1}^N$

PINN

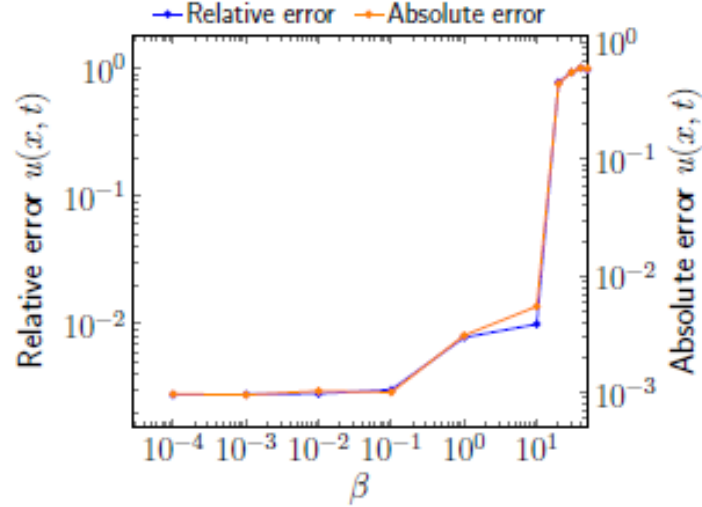
- Anahtar nokta : Kısmi türevler NN yapısından AD ile hesaplanıyor
- Orijinal formülasyon : soft-constraints

$$MSE := \frac{1}{N} \sum_{i=1}^N \left(|u(t^i, x^i, y^i) - u^i|^2 + |v(t^i, x^i, y^i) - v^i|^2 \right) + \frac{1}{N} \sum_{i=1}^N \left(|f(t^i, x^i, y^i)|^2 + |g(t^i, x^i, y^i)|^2 \right).$$



Correct PDE	$u_t + (uu_x + vu_y) = -p_x + 0.01(u_{xx} + u_{yy})$ $v_t + (uv_x + vv_y) = -p_y + 0.01(v_{xx} + v_{yy})$
Identified PDE (clean data)	$u_t + 0.999(uu_x + vu_y) = -p_x + 0.01047(u_{xx} + u_{yy})$ $v_t + 0.999(uv_x + vv_y) = -p_y + 0.01047(v_{xx} + v_{yy})$
Identified PDE (1% noise)	$u_t + 0.998(uu_x + vu_y) = -p_x + 0.01057(u_{xx} + u_{yy})$ $v_t + 0.998(uv_x + vv_y) = -p_y + 0.01057(v_{xx} + v_{yy})$

Prediction error for 1D convection (§3.1) problem, when β is changed.



(a) Error for different β

Characterizing possible failure modes in physics-informed neural networks

Aditi S. Krishnapriyan^{*,1,2}, Amir Gholami^{*,2},
Shandian Zhe³, Robert M. Kirby³, Michael W. Mahoney^{2,4}

¹Lawrence Berkeley National Laboratory, ²University of California, Berkeley,

³University of Utah, ⁴International Computer Science Institute

(2021)

- “Soft-constraint” formülasyonu problemin sağlamlığını kötüleştiriyor (ill-conditioned)
 - Sadece küçük katsayıları olan kolay problemler için iyi çalışıyor, azıcık zor fiziksel rejimler için çok büyük hatalar üretebiliyor
 - Hataların kaynağı ‘soft-constraint’ formülasyonu; aslında NN yapısı iyi çözüm bulma kapasitesine sahip
- İki öneri sunulmuş
 - Problemi önce küçük katsayılar için çözüp, yavaş yavaş katsayıyı büyütme (curriculum regularization)
 - Problemi bütün uzay-zaman için tek seferde çözmek yerine, bir seferde tek bir zaman aralığı için çözmek (seq2seq learning)

Tek seviye kısıtlı stokastik optimizasyon problemi

- Amaç fonksiyonu ve kısıt fonksiyonları **belirsizliği** modelleyen rassal parametrelere bağlı
 - Ölçüm verilerindeki hatalar
 - Proses parametrelerine ilişkin eksik bilgi

$$\begin{array}{ll}\min & f(x, \theta) \\ \text{s.t.} & c(x, \theta) \leq 0.\end{array}$$



$$\begin{array}{ll}\min & \mathcal{R}^f(x) \\ \text{s.t.} & \mathcal{R}^c(x) \leq 0.\end{array}$$

- Sağlamlık (robustness) ölçütleri : \mathcal{R}^f ve \mathcal{R}^c – sadece yaklaşık olarak hesaplayabiliyoruz
 - Beklenen değer ve varyansın kombinasyonları, en kötü durum, olasılıksal kısıtlar (value at risk)

A trust-region method for derivative-free nonlinear constrained stochastic optimization

Kısıtlı stokastik optimizasyon

- Stokastik optimizasyon büyük ve uzun yıllardır çalışılan bir alan
 - *SAA (Sample average approximation) : Yaklaşık fonksiyonu (örneğin yeterince büyük bir örnek seçerek) optimizasyon süreci boyunca koru*
 - Birden fazla run alınarak ortalaması hesaplanabilir ya da örnek büyüklüğü arttırılabilir
 - *SA (Stokastik approximation) : Her fonksiyon hesaplamasında yeni bir örnek seç ve örnekleme hatası giderek azalmak zorunda değil*
 - En büyük dezavantaj adım boyu gibi parametrelerin performansa etkisinin büyük olması ve aynı zamanda pratikte nasıl seçileceklerinin net olmaması / probleme bağlı olması
 - *Diğer alternatifler*
 - Constrained Bayesian optimization (GP)
 - Constrained DFO (eğer örnekleme hatası küçükse ya da giderek azalıyorsa)

Kısıtlı stokastik optimizasyon

- Deterministik kısıtlı optimizasyon yöntemleri ve varyantlarını SA yaklaşımına adapte eden bir dizi yeni çalışma var
 - *Stokastik amaç fonksiyonu, deterministik (eşitlik) kısıtlar*

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad c(x) = 0, \quad \text{with} \quad f(x) = \mathbb{E}[F(x, \iota)]$$

Sequential Quadratic Optimization for Nonlinear Equality
Constrained Stochastic Optimization

(2022)

ALBERT BERAHAS^{*1}, FRANK E. CURTIS^{†2}, DANIEL P. ROBINSON^{‡2}
BAOYU ZHOU^{§2}

(2020)

An Adaptive Stochastic Sequential Quadratic
Programming with Differentiable Exact Augmented
Lagrangians

Sen Na • Mihai Anitescu • Mladen Kolar

Fully Stochastic Trust-Region Sequential Quadratic Programming
for Equality-Constrained Optimization Problems

(2022)

Yuchen Fang¹, Sen Na², Michael W. Mahoney^{2,3}, and Mladen Kolar⁴

Entegrasyon

- Öğrenilen fonksiyon başka bir matematiksel modelin bileşeni olarak kullanılabilir
 - **İki aşama** : Önce bileşen fonksiyonları öğren, sonra ana modeli çöz
 - **Tek aşama** : Optimallik koşullarını öğrenme problemine entegre et

With \mathbf{y}_*^i denoting the training data for the input (t^i, x^i) , and $\mathbf{y}^i, \mathbf{u}^i, \boldsymbol{\lambda}^i$ corresponding outputs of the network, $\boldsymbol{\theta}_k$ representing the weights of the network at iteration k , and $\|\cdot\|$ as L_2 norm, we seek to minimize the loss function:

“Physics-informed neural networks for
PDE-constrained optimization and control”

Jostein Barry-Straume, Arash Sarshar,
Andrey A. Popov, and Adrian Sandu

Computational Science Laboratory
“Compute the Future!”

Department of Computer Science
Virginia Tech

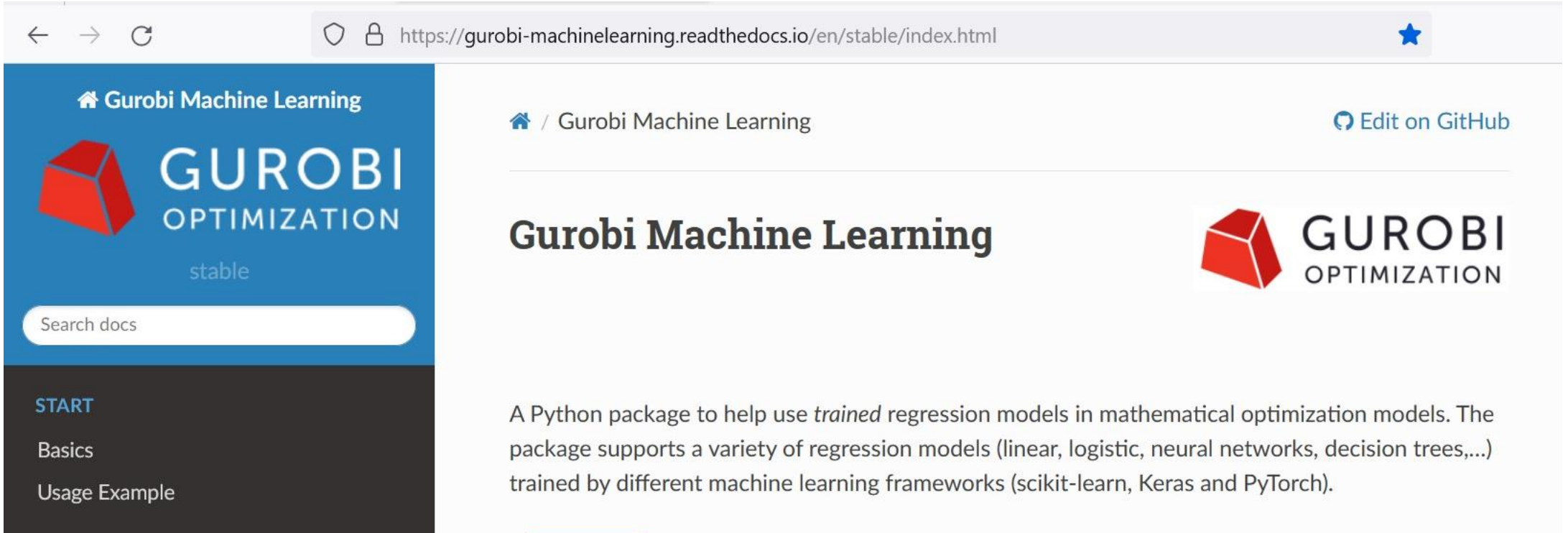
$$L(\boldsymbol{\theta}_k) = \sum_i \|\mathbf{y}_*^i - \mathbf{y}^i\|^2 + \sum_i \left\| \frac{\partial \mathbf{y}^i}{\partial t} - f(\mathbf{y}^i, \mathbf{u}^i) \right\|^2 + \sum_i \left\| \frac{\partial \boldsymbol{\lambda}^i}{\partial t} + (\boldsymbol{\lambda}^i)^T \frac{\partial \mathbf{f}}{\partial \mathbf{y}^i}(\mathbf{y}^i, \mathbf{u}^i) + \frac{\partial g}{\partial \mathbf{y}^i}(\mathbf{y}^i, \mathbf{u}^i) \right\|^2 + \sum_i \left\| (\boldsymbol{\lambda}^i)^T \frac{\partial \mathbf{f}}{\partial \mathbf{u}^i}(\mathbf{y}^i, \mathbf{u}^i) + \frac{\partial g}{\partial \mathbf{u}^i}(\mathbf{y}^i, \mathbf{u}^i) \right\|^2.$$

*PDE-kısıtlı
problem için
optimallik
koşulları*

(1)

Öğrenilmiş fonksiyonlarla optimizasyon

- Öğrenilmiş fonksiyonları optimizasyon modellerinde kullanmak sadece akademik bir fikir değil



The screenshot shows a web browser displaying the Gurobi Machine Learning documentation. The browser's address bar shows the URL <https://gurobi-machinelearning.readthedocs.io/en/stable/index.html>. The page features a blue header with the Gurobi logo and the text "GUROBI OPTIMIZATION stable". A search bar labeled "Search docs" is present. On the left, a dark sidebar lists navigation options: "START", "Basics", and "Usage Example". The main content area has a light blue header with a home icon and the text "Gurobi Machine Learning", and a link to "Edit on GitHub". The main heading is "Gurobi Machine Learning", followed by the Gurobi logo and the text "GUROBI OPTIMIZATION". The introductory text states: "A Python package to help use *trained* regression models in mathematical optimization models. The package supports a variety of regression models (linear, logistic, neural networks, decision trees,...) trained by different machine learning frameworks (scikit-learn, Keras and PyTorch)."

Öğrenilmiş fonksiyonlarla optimizasyon

- Gurobi ML, regresyon, yapay sınır ağı, karar ağacı, vb ile öğrenilen modelleri karışık-tam sayılı programlama modellerine dönüştürerek ana çözücüyeye aktarıyor
 - Örneğin, lojistik regresyon için parçalı doğrusal yaklaşık fonksiyon
 - Yapay sinir ağı için yalnızca ReLU aktivasyon fonksiyonu ile çalışabiliyor

$$\max_x \quad \sum_{j=1}^{n_1} c_j x_j + \sum_{k=1}^{n_2} d_k y_k$$

$$\text{s.t.} \quad \sum_{j=1}^{n_1} a_j^i x_j \leq b_i,$$

e_l^k : binary unit vector

$$y_k = g_k(\alpha_1^k, \dots, \alpha_{p_k}^k; \theta_k),$$

$$\alpha_l^k \text{ is given,}$$

$$\alpha_l^k = e_l^k \cdot x,$$

$$x_j \in X_j,$$

JANOS : An Integrated Predictive and Prescriptive Modeling Framework

David Bergman¹, Teng Huang¹, Philip Brooks², Andrea Lodi³, and Arvind U. Raghunathan⁴

$$\forall i \in \{1, \dots, m\} \quad (1)$$

$$\forall k \in \{1, \dots, n_2\} \quad (2)$$

$$\forall l \in \{1, \dots, q_k\}, \forall k \in \{1, \dots, n_2\} \quad (3)$$

$$\forall l \in \{(q_k + 1), \dots, p_k\}, \forall k \in \{1, \dots, n_2\} \quad (4)$$

$$\forall j \in \{1, \dots, n_1\}. \quad (5)$$

Örnek : Tasarım optimizasyonu

Modeling Design and Control Problems Involving Neural Network Surrogates

Dominic Yang · Prasanna Balaprakash · Sven Leyffer

$$\begin{array}{ll} \underset{f, r, c, \text{CO}, \text{NO}, \text{Torque}}{\text{minimize}} & \sum_{t=1}^T (\text{NO}_t + \lambda \text{CO}_t) \Delta t & \text{min. NO and CO emissions} \\ \text{subject to} & \text{Torque}_t \geq \text{PrescribedTorque}_t, \forall t = 1, \dots, T & \text{trip profile} \\ & (\text{NO}_t, \text{CO}_t, \text{Torque}_t) = \text{DNN}(f_t, r_t, c) & \text{DNN constraints} \\ & f_t \in [f_{\min}, f_{\max}], r_t \in [r_{\min}, r_{\max}], c \in [c_{\min}, c_{\max}] & \text{bounds on controls.} \end{array}$$

- Otomobil motor spesifikasyonlarını emisyon miktarı ve tork ile ilişkilendiren NN modeli
 - Her zaman aralığı t için farklı DNN – problem 1500 zaman aralığı için çözülüyor (25 dakikalık sürüşteki 1500 saniye)
 - Tanım kümesinin öğrenilen alanının kullanıyor olmak kritik
 - *Eğer eğitim veri setinin konveks zarfı (convex hull) içinde olma kısıtı eklenmezse optimal tasarım negatif emisyon değerleri üretebiliyor.*
 - Üç çözüm yaklaşımı denenmiş ReLU için: NLP (embedded), MPCC, MIP
 - MIP formülasyonu optimal çözümü sağlamakla birlikte çözüm süresi açısından ölçeklenebilir görünmüyor

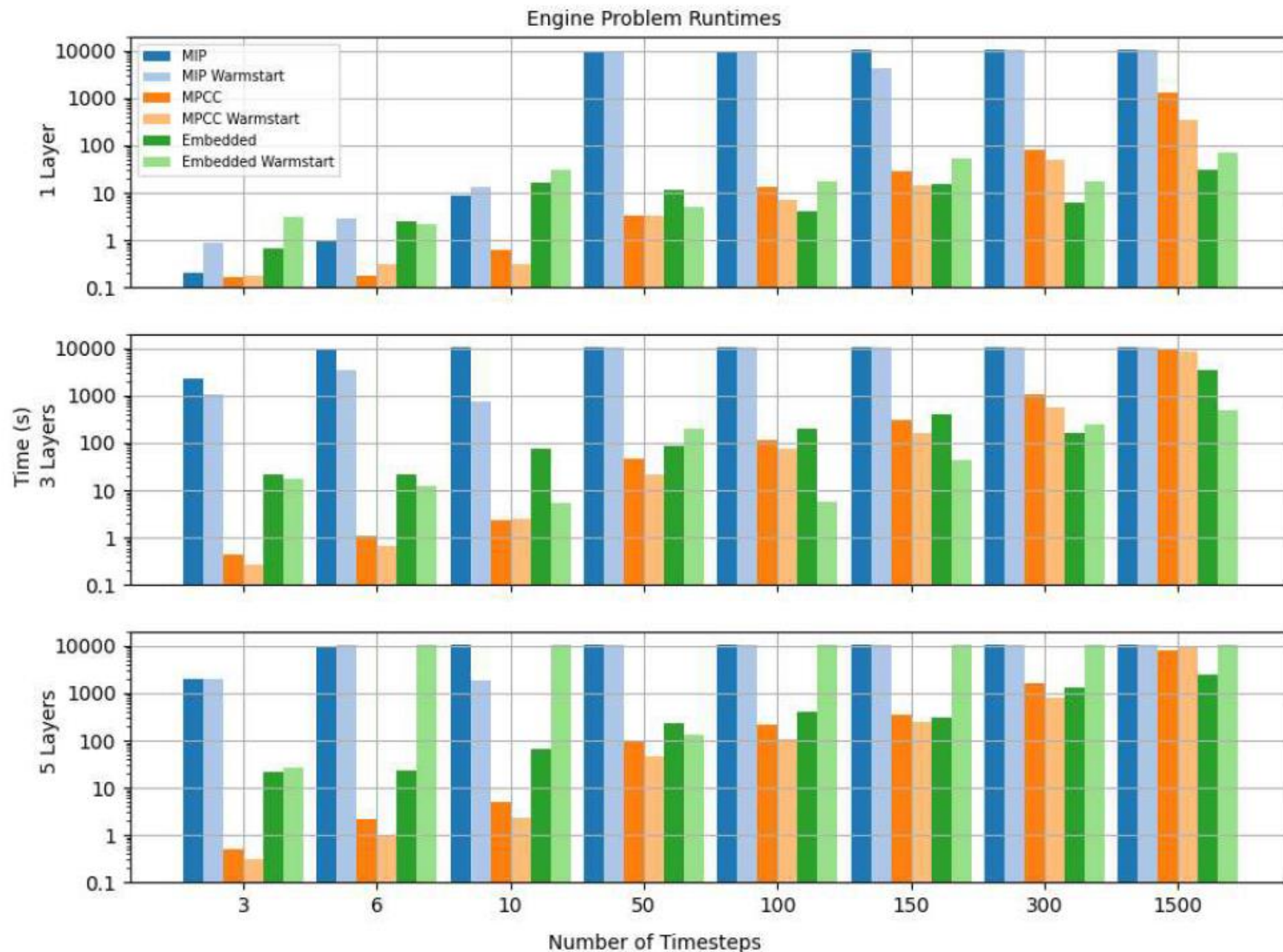


Fig. 8: Time until a solver found its best solution within 3 hours (10,800 seconds) on each architecture for varying values of T . If a runtime of 10,800 seconds is recorded, this indicates the solver failed to find a feasible solution in the full 3 hours (if a warmstart is used, this means it failed to find a feasible solution better than the warmstart).

Gürültülü fonksiyon hesaplamaları ile kısıtlı optimizasyon

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$c_E(x) = 0$$

$$c_I(x) \leq 0$$

$$\tilde{f}(x) = f(x) + \epsilon_f(x),$$

$$\tilde{g}(x) = \nabla f(x) + \epsilon_g(x),$$

$$\tilde{c}(x) = c(x) + \epsilon_c(x),$$

$$\tilde{J}(x) = \nabla c(x) + \epsilon_J(x).$$

Constrained Optimization in the Presence of Noise

Figen Oztoprak*

Richard Byrd[†]

Jorge Nocedal[‡]

- Amaç ve kısıt fonksiyonları ile bunların türevlerinin sadece gürültülü (noisy) değerlerine erişebiliyoruz
 - *Bu en genel durum; gürültü bileşenlerinden bazıları sıfır olabilir (örnek : DFO)*
- Buradaki gürültü, **stokastik** ya da deterministik olabilir, fakat sabit değil

Potansiyel araştırma soruları

- Kısıtlı stokastik optimizasyon
 - *Makine öğrenmesi alanındaki gerçek pratik potansiyel*
 - Yakın dönem optimizasyon-komünitesi makalelerindeki test problemleri yapay
 - *Öğrenmenin kısıtlı optimizasyon problemi olarak formülasyonu (kısıtsız formülasyona indirgemek yerine) durumunda etkin çözümü*
 - Örneğin PINN için
- Öğrenilen kısıtlarla optimizasyon
 - *Modeli lokal olarak öğrenmek*
 - Çözüm algoritması ilerlerken ‘güncellenen’ kısıt fonksiyon modelleri ile optimizasyon için hangi yöntem(ler) nasıl çözümler verebilir? Stabilite garantileri verilebilir mi?