Optimization with Learning-Informed Differential Equation Constraints

Michael Hintermüller[1,2]

[1]Weierstrass Institute for Applied Analysis and Stochastics (WIAS),
[2]Humboldt-Universität zu Berlin

EMFCSC School G. Stampacchia

Joint work with Guozhi Dong and Kostas Papafitsoros

May 24, 2022

Data-driven methods for model-based prediction

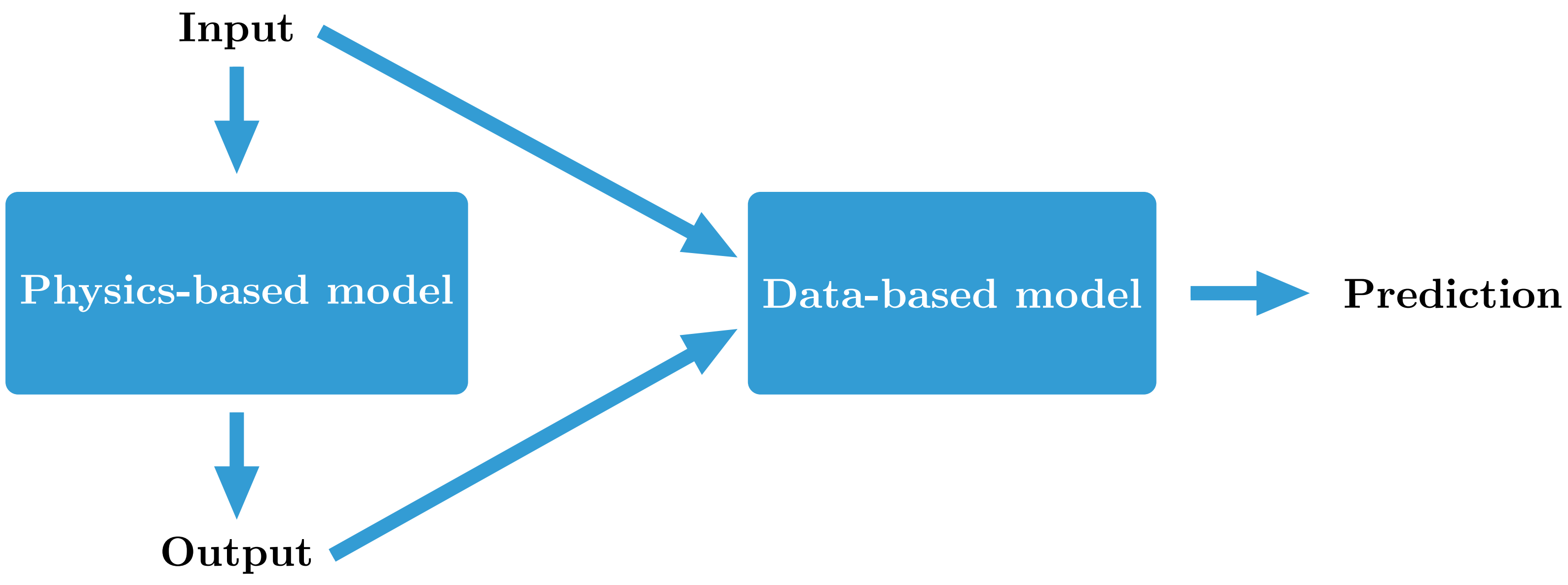


Figure: Ab initio models typically used to analyze experimental data and for prediction

- Making the physics model more and more accurate is a continuous challenge.
- Artificial neural networks are efficient tools to learn physical laws from data.
- Taking advantage of ever increasing computational power and data availability.

A general optimization workflow with learned physics

Learning-informed models as constraints in optimization

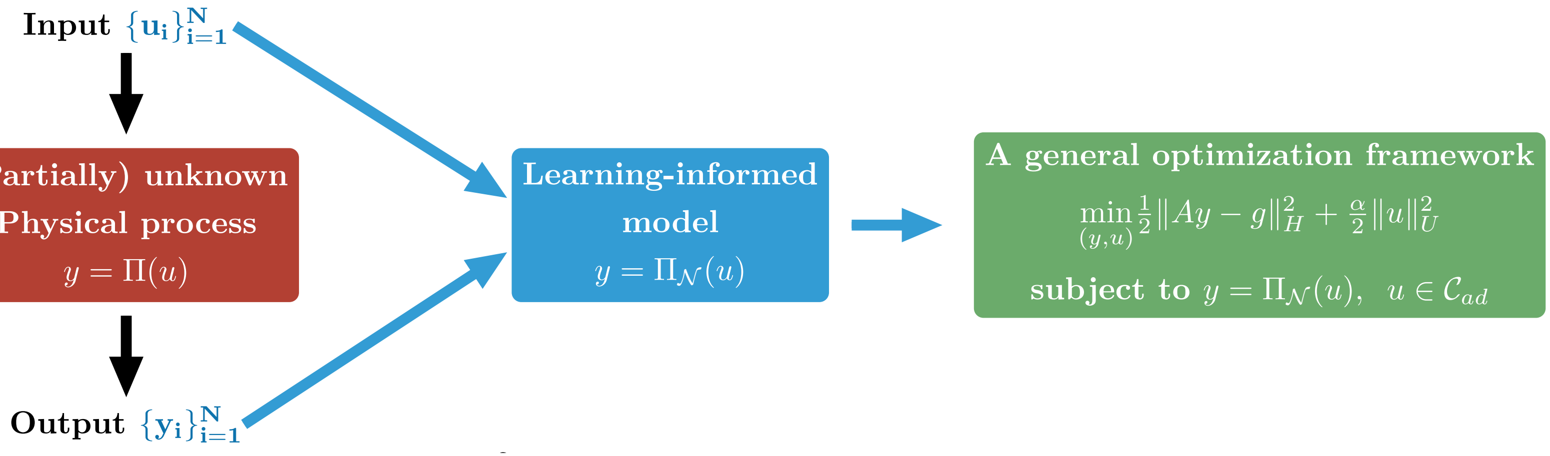
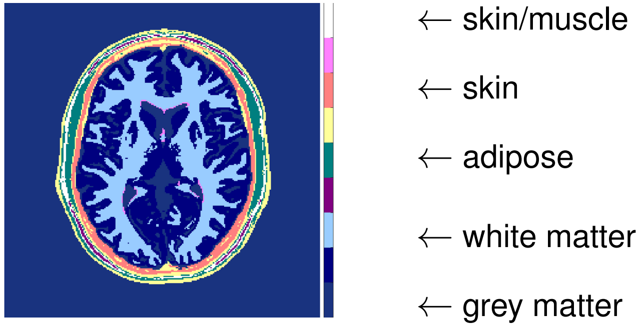


Figure: Workflow of optimization with learning-informed physical constraints

---

Phase separation

Quantitative MRI



Contents

Mathematics of deep learning and its "current state"

Artificial neural networks (ANNs) in brief

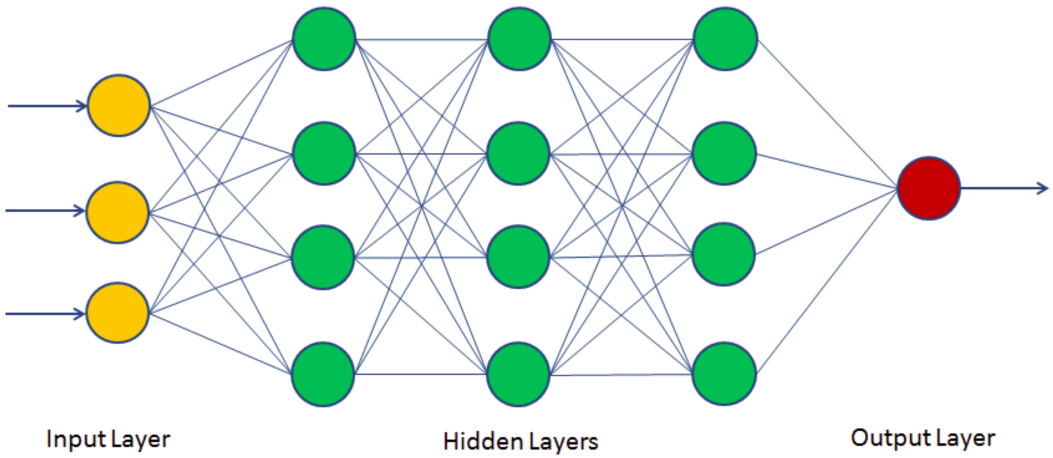


Figure: A diagram of an artificial neural network

Key components:

- u: input data
- y: output data
- $h^{(l+1)} = \sigma(W_l h^l + b_l)$
- σ: activation function
- W_l: weight matrix
- b_l: bias vector
- One hidden-layer case: $\mathcal{N}(u) := W_1\sigma(W_0 u + b_0) + b_1 \to y$
- W_l and b_l are unknowns to be fixed

Supervised learning is about solving the following generic optimization problem:

$$\underset{(W,b)\in\mathcal{F}_{ad}}{\text{minimize}} \quad \sum_{j=1}^{n} \mathfrak{d}(\mathcal{N}(u_j), y_j) + \mathfrak{r}(W, b)$$

for given training data pairs $(u_j, y_j)_{j=1}^n$, and $W := (W_l)_{l=0}^L$, $b := (b_l)_{l=0}^L$.

Universal approximation theorem[1]

ANNs have been very successful approximators for functions $f : \Omega \to \mathbb{R}^n$, defined on bounded $\Omega \subset \mathbb{R}^m$.

Theorem (function value approximation)

A standard multi-layer feedforward network with a continuous activation function can uniformly approximate any continuous function to any degree of accuracy if and only if its activation function is not a polynomial.

Theorem (derivative approximation)

There exists a neural network which can approximate both the function value and the derivatives of f uniformly to any degree of accuracy if the activation function is continuously differentiable and is not a polynomial.

[1] Pinkus, Approximation theory of the MLP model in neural networks. Acta Numerica, 1999.

Activation functions of ANNs

Examples of smooth activation functions:

- Sigmoid: e.g., tansig ($\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$), logsig ($\sigma(z) = \frac{1}{1+e^{-z}}$)), arctan ($\sigma(z) = \arctan(z)$)), etc.
- Probability functions: e.g., softmax ($\sigma_i(z) = \frac{e^{-z_i}}{\sum_j e^{-z_j}}$)

Examples of nonsmooth activation functions:

- ReLU: Rectified Linear Unit ($\sigma(z) = \max(0, z)$)

Important: Choosing smooth vs. nonsmooth activation functions should respect prior information on to be approximated object and has numerous implications in optimization.

Current state on ANN's approximation

NNs approximate an objective f in different settings	Examples
1. $f : \Omega \subset \mathbb{R}^m \to \mathbb{R}^n$, with finite m and n Universal approximation theorem	▪ (Generalized) Regression
2. $f : \mathcal{K} \subset \mathcal{B}_1 \to \mathbb{R}^n$, where $\mathcal{B}_1$ is some Banach space Under-development (mostly convolutionary NNs)	▪ (Image) Classification
3. $f : \Omega \subset \mathbb{R}^m \to \mathcal{B}_2$, where $\mathcal{B}_2$ is some Banach space Under-development (many different methods)	▪ Solving (partial) differential equations
4. $f : \mathcal{K} \subset \mathcal{B}_1 \to \mathcal{B}_2$, $(\mathcal{B}_k)_{k=1}^2$ can be infinite dimensional Under-development (very few still)	▪ Operator learning

Except for case 1, mathematical understanding of cases 2–4 still mostly in progress.

Main difficulty: Compactness condition problematic.

Physics-informed learning[2] vs Learning-informed physics[3]

Physics-informed learning

- Physical models enter learning and neural networks
- PDE residuals are part of loss function for training
- Usually of type $f : \Omega \to \mathcal{B}_2$

Learning-informed physics

- Using ANNs to predict physical models or their constituents
- Loss function is not necessarily PDE dependent
- Typically of type $f : \mathcal{B}_1 \to \mathcal{B}_2$

To directly learn operators between Banach spaces using ANNs has been intensively investigated recently [a,b].

[a] Bhattacharya, Hosseini, Kovachki and Stuart, Model reduction and neural networks for parametric PDEs, ICLR, 2021.

[b] Lu, Jin, Pang, Zhang & Karniadakis, Learning nonlinear. operators via DeepONet based on the universal approximation theorem of operators, Nature Machine Intell., 2021.

[2] Rassi, Perdikaris and Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear PDEs. J. Comp. Phys. 2019.

[3] Dong, Hintermüller and Papafitsoros, Optimization with learning-informed differential equation constraints and its applications, to appear in ESAIM: COCV, 2021.

Optimization constrained by learning-informed models

A general framework involving physics-based models

We study the following optimization problem:

$$
\begin{aligned}
\underset{(y,u)\in(Y\times U)}{\text{minimize}} \quad & \frac{1}{2}\|Ay-g\|_H^2+\frac{\alpha}{2}\|u\|_U^2, \\
\text{subject to} \quad & e(y,u)=0, \\
& u\in\mathcal{C}_{ad}.
\end{aligned}
$$

- $A : U \to Y$ a bounded, linear operator
- $e(y,u) = 0$ physical model; e.g., (system of) ODEs or PDEs

A general framework involving physics-based models

We study the following optimization problem:

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & \frac{1}{2}\|A\Pi(u) - g\|_H^2 + \frac{\alpha}{2}\|u\|_U^2 =: \mathcal{J}(u), \\ \text{subject to} \quad & u \in \mathcal{C}_{ad}. \end{aligned}$$

- $A : U \to Y$ a bounded, linear operator
- $e(y, u) = 0$ physical model; e.g., (system of) ODEs or PDEs
- Well-posedness $e(y, u) = 0$ leads to $y = \Pi(u)$

A general framework involving physics-based models

We study the following optimization problem:

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & \frac{1}{2}\|A\Pi_{\mathcal{N}}(u) - g\|_H^2 + \frac{\alpha}{2}\|u\|_U^2 =: \mathcal{J}_{\mathcal{N}}(u), \\ \text{subject to} \quad & u \in \mathcal{C}_{ad}. \end{aligned}$$

- $A : U \to Y$ a bounded, linear operator
- $e(y, u) = 0$ physical model; e.g., (system of) ODEs or PDEs
- Well-posedness $e(y, u) = 0$ leads to $y = \Pi(u)$
- ANNs for operator learning yield $\Pi_{\mathcal{N}} \sim \Pi$ (possibly via different pathways)

A general framework involving physics-based models

We study the following optimization problem:

$$\underset{u}{\text{minimize}} \quad \frac{1}{2}\|A\Pi_{\mathcal{N}}(u) - g\|_H^2 + \frac{\alpha}{2}\|u\|_U^2 =: \mathcal{J}_{\mathcal{N}}(u),$$
$$\text{subject to} \quad u \in \mathcal{C}_{ad}.$$

- $A : U \to Y$ a bounded, linear operator
- $e(y, u) = 0$ physical model; e.g., (system of) ODEs or PDEs
- Well-posedness $e(y, u) = 0$ leads to $y = \Pi(u)$
- ANNs for operator learning yield $\Pi_{\mathcal{N}} \sim \Pi$ (possibly via different pathways)

Fundamental questions:

- Conditions for well-posedness of learned physical model and universal approximation property of $\Pi_{\mathcal{N}} \sim \Pi$.
- Approximation properties of optimizers associated to learning-informed models vs. those related to original physics-based models.

Let $Q := A\Pi$ (or $A\Pi_{\mathcal{N}}$).

Proposition

Suppose that Q is weakly-weakly sequentially closed, i.e., if $u_n \overset{U}{\rightharpoonup} u$ and $Q(u_n) \overset{H}{\rightharpoonup} \bar{g}$, then $\bar{g} = Q(u)$. Then the optimization problem admits a solution $\bar{u} \in U$.
In the special case when $\mathcal{C}_{ad}$ is a bounded set of a subspace $\hat{U}$ compactly embedded into U, then strong-weak sequential closedness of Q is sufficient to guarantee existence of a solution.

- In many PDE models, regularity of the resp. solution helps the weak-weak sequential closedness condition of the control-to-state map to be satisfied.
- While in imaging applications (inverse problems, more generally) regularization usually plays a role similar to $\hat{U}$.

Convergence under operator perturbations

Let $Q_n := A\Pi_{\mathcal{N}_n}$ be the reduced learning-informed operators.

Theorem

Let Q and Q_n for $n \in \mathbb{N}$ be weakly sequentially closed operators, and

$$\sup_{u \in \mathcal{C}_{ad}} \|Q(u) - Q_n(u)\|_H \leq \epsilon_n, \quad \textit{for} \quad \epsilon_n \downarrow 0.$$

Suppose $(u_n)_{n\in\mathbb{N}}$ is a sequence of minimizers associated to the optimization problems with reduced operator Q_n for all $n \in \mathbb{N}$.
Then, there is the strong convergence (up to a sub-sequence)

$$u_n \to \bar{u} \ \textit{ in } \ U, \quad \textit{and} \quad Q_n(u_n) \to Q(\bar{u}) \ \textit{ in } \ H, \quad \textit{as} \ n \to \infty,$$

where $\bar{u}$ is a minimizer of the original optimization problem.

Convergence rates

Denote L_0 and L_1 the Lipschitz constants associated to Q and Q', respectively, where Q' is the Fréchet derivative of Q, and $\eta_n := \|Q' - Q'_n\|_{\mathcal{L}(U,H)}$.

Theorem

Under smallness of L_0, L_1, the solutions u_n converge to $\bar{u}$ at the following rate

$$\|u_n - \bar{u}\|_U = \mathcal{O}\left(L_0 \epsilon_n + \|Q(\bar{u}) - g\|_H \eta_n\right).$$

Theorem (when $\mathcal{J}'(\bar{u}) = 0$)

Suppose the Lipschitz constant L_1 satisfies

$$L_1 \|Q(\bar{u}) - g\|_H < \alpha.$$

If $\mathcal{J}'(\bar{u}) = 0$, then for sufficiently large $n \in \mathbb{N}$ we have the following error bound

$$\|u_n - \bar{u}\|_U = \mathcal{O}\left(\sqrt{\epsilon_n^2 + 2\|Q(\bar{u}) - g\|_H^2}\right).$$

Case studies

Learn control-to-state map for semilinear PDEs

We consider the following model problem:

$$\begin{aligned} &\underset{(y,u)}{\text{minimize}} \quad \frac{1}{2}\|y-g\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}, \\ &\text{subject to } -\Delta y + f(\cdot, y) = u \quad \text{in } \Omega, \quad \partial_\nu y = 0 \text{ on } \partial\Omega, \\ &\qquad u \in \mathcal{C}_{ad} := \{v \in L^2(\Omega) : \underline{u}(x) \le v(x) \le \overline{u}(x), \quad \text{for } x \in \Omega\}. \end{aligned}$$

- f is some unknown map, e.g., modeling phase separation
- Goal is to learn the control-to-state (C2S) map $\Pi : u \to y$

Learn control-to-state map for semilinear PDEs

We consider the following model problem:

$$\begin{aligned} &\underset{(y,u)}{\text{minimize}} \quad \frac{1}{2}\|y-g\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}, \\ &\text{subject to } -\Delta y + \mathcal{N}(\cdot, y) = u \quad \text{in } \Omega, \quad \partial_\nu y = 0 \text{ on } \partial\Omega, \\ &\qquad u \in \mathcal{C}_{ad} := \{v \in L^2(\Omega) : \underline{u}(x) \le v(x) \le \overline{u}(x), \quad \text{for } x \in \Omega\}. \end{aligned}$$

- f is some unknown map, e.g., modeling phase separation
- Goal is to learn the control-to-state (C2S) map $\Pi : u \to y$
- Ideal: learn f through a neural network $\mathcal{N}$ via $f(\cdot, y) = \Delta y + u$
- The learning-informed PDE with component $\mathcal{N}$, induces the C2S map $\Pi_{\mathcal{N}}$

Assumptions on the nonlinearity

- (Regularity) $f = f(x,z) : \Omega \times \mathbb{R} \to \mathbb{R}$ is measurable in x and continuous in z.
- (Growth-rate) There is $F : \Omega \times \mathbb{R} \to \mathbb{R}$ so that $\partial_z F(\cdot, z) = f(\cdot, z)$, satisfying

$$|f(\cdot,z)| \leq b_1 + c_1 |z|^{p-1} \quad \text{and} \quad -f(\cdot,z)z + F(\cdot,z) \leq b_2,$$

  resulting in

$$F(\cdot,z) \leq b_0 + c_0 |z|^p ,$$

  for some constants $b_0, b_1, b_2 \in \mathbb{R}$ and $c_0, c_1 > 0$, and for some p so that the embedding $H^1(\Omega) \subset L^p(\Omega)$ holds.
- (Coercivity) F is coercive in the sense that $\lim_{\|y\|_{L^p(\Omega)} \to \infty} \frac{\int_\Omega F(x,y)dx}{\|y\|_{L^p(\Omega)}} = \infty$.
- (Boundedness) F is bounded from below, i.e., $F(x,z) \geq F_0$ for some $F_0 \in \mathbb{R}$.

A priori bounds on PDE solutions

A variational problem connected to nonlinear PDE:

$$\inf G(y) := \frac{1}{2}\|\nabla y\|^2_{L^2(\Omega)} + \int_\Omega F(x,y)\,dx - \int_\Omega uy\,dx \quad \text{over } \ y \in H^1(\Omega). \tag{3.1}$$

Proposition

Suppose that $u \in L^r(\Omega)$ for some $r \geq \frac{p}{p-1}$. Then the optimization problem (3.1) *admits a solution in $H^1(\Omega)$, which satisfies the constraint PDE.*

Theorem

Let $\mathcal{C}_{ad} \subset L^\infty(\Omega)$ be bounded. Then there exists a constant $K > 0$ such that for all solutions y of the semilinear PDE, it holds

$$\|y\|_{H^1(\Omega)} + \|y\|_{C(\overline{\Omega})} \leq K, \quad \textit{for all } \ u \in \mathcal{C}_{ad}.$$

Existence of solutions for learning-informed PDEs

Proposition

Let $f : \Omega \times \mathbb{R} \to \mathbb{R}$ and $F : \Omega \times \mathbb{R} \to \mathbb{R}$ be given as before with the extra assumption that $f \in C(\overline{\Omega} \times \mathbb{R})$. Then, for every $\epsilon > 0$ there exists a neural network $\mathcal{N} \in C^\infty(\mathbb{R}^d \times \mathbb{R})$ such that

$$\sup_{\|y\|_{L^\infty(\Omega)} < K} \|f(\cdot, y) - \mathcal{N}(\cdot, y)\|_U < \epsilon, \tag{3.2}$$

with K the uniform bound. Moreover, the learning-informed PDE

$$-\Delta y + \mathcal{N}(\cdot, y) = u \quad \textit{in } \Omega, \qquad \partial_\nu y = 0 \ \textit{ on } \partial\Omega,$$

admits a weak solution which satisfies the a priori bound for sufficiently small $\epsilon > 0$.

Only approximation property $\|y\|_{L^\infty(\Omega)} < K$ is needed in (3.2) .

Theorem (under constraint on negative part of $\partial_y f(\cdot, y_0)$)

Suppose $u_n = u_0 + t_n h$ for a sequence $t_n \to 0$, and suppose there exists $y_n \in \Pi_{\mathcal{N}}(u_n)$ with $y_n \to y_0$ in $H^1(\Omega)$. Then, we have

- *Local Lipschitz property:*

$$\|y_n - y_0\|_{H^1(\Omega)} \leq C t_n,$$

  for some constant C.

- *Directional differentiability: Every weak cluster point of $\frac{y_n - y_0}{t_n}$, denoted by p, solves*

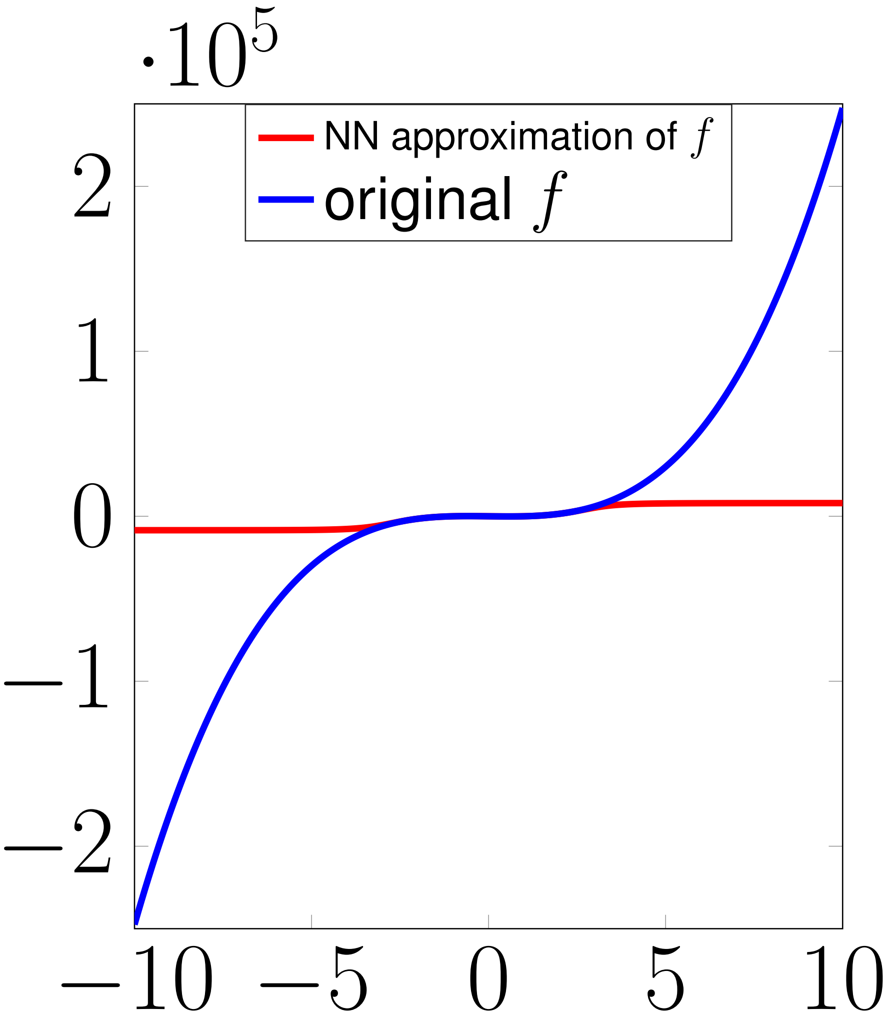
$$-\Delta p + \partial_y \mathcal{N}(\cdot, y_0) p = h \quad \text{in } \Omega, \qquad \partial_\nu p = 0 \text{ on } \partial\Omega,$$

  and p satisfies the energy bounds for every $h \in L^2(\Omega)$,

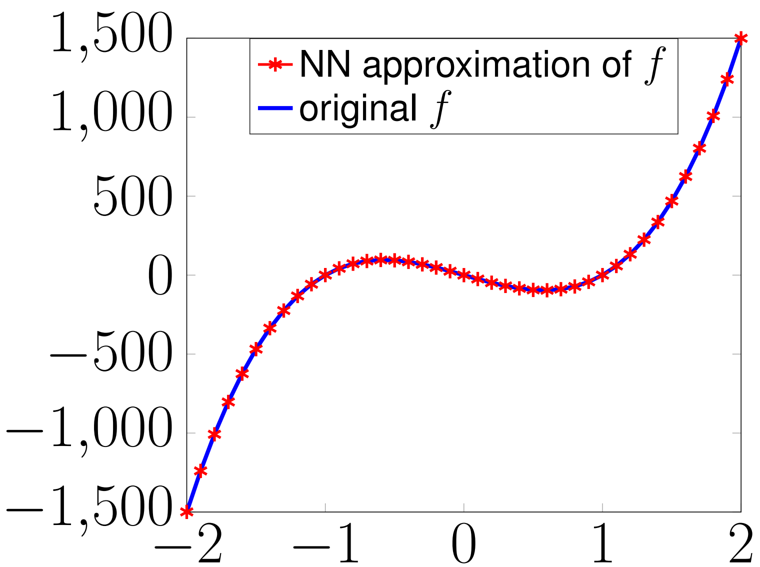
$$\|p\|_{H^1(\Omega) \cap C(\overline{\Omega})} \leq C \|h\|_{L^2(\Omega)}$$

  for some constant C.

Learning-informed double-well potential

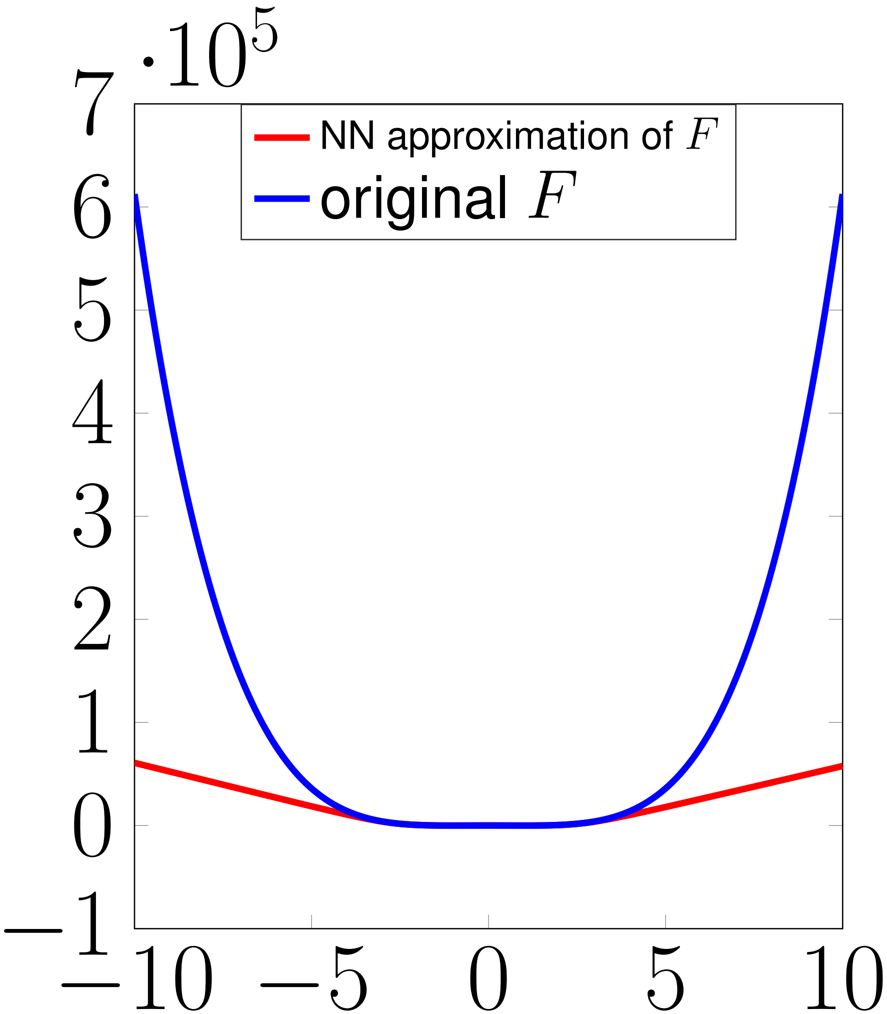


Approximation of $f(y) = \frac{1}{0.004}(y^3 - y)$ by neural network function.

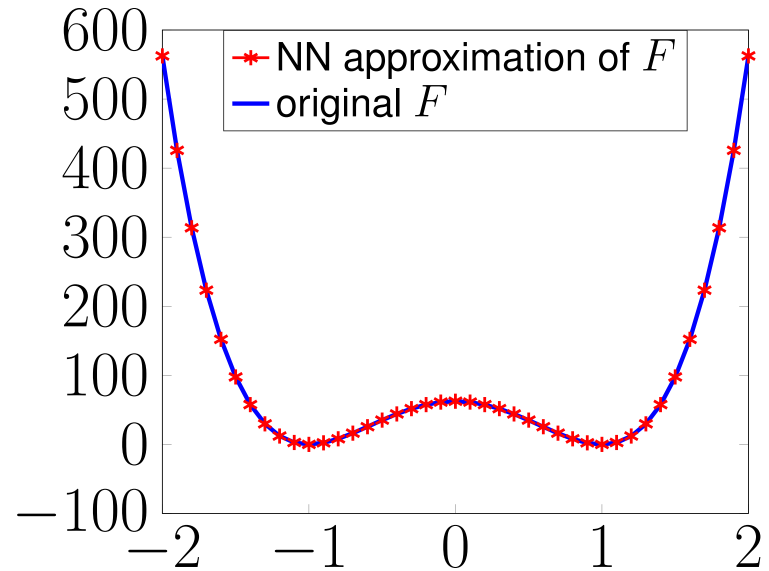


Details of the left Figure

Learning-informed double-well potential



The double well potential F and $F_{\mathcal{N}}$ reconstructed from f and $\mathcal{N}$, respectively.



Details of the left Figure

Universal approximation of learning-informed operator

Proposition

There exists $\mathcal{N} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ *so that*

$$\sup_{\|y\|_{L^\infty(\Omega)} < M} \|f(\cdot, y) - \mathcal{N}(\cdot, y)\|_U \leq \epsilon,$$

for $\epsilon > 0$ *arbitrarily small. Further, we have the error bounds*

$$\|\Pi(u) - \Pi_{\mathcal{N}}(u)\|_H \leq C\epsilon, \quad \textit{for all } u \in \mathcal{C}_{ad},$$

where the constant $C > 0$ *depends on* f *and* y_0*. When* f *is locally Lipschitz, there exists also* $\mathcal{N}$ *so that*

$$\sup_{\|y\|_{L^\infty(\Omega)} < M} \|\partial_y f(\cdot, y) - \partial_y \mathcal{N}(\cdot, y)\|_U \leq \epsilon_1,$$

for sufficiently small $\epsilon_1 > 0$*, and there exist some constants* $C_0 > 0$ *and* $C_1 > 0$

$$\|p_0 - p_\epsilon\|_{H^1(\Omega)\cap C(\overline{\Omega})} \leq C_1\epsilon_1 + C_0\epsilon, \quad \textit{for all } u \in \mathcal{C}_{ad}.$$

The adjoint variables p_ϵ, p_0 are directional derivatives of $\Pi_{\mathcal{N}}$ and Π, respectively.

KKT condition and semismooth Newton method

The KKT system of the optimal control problem $(A = I, \mathcal{C}_{\text{ad}}$ a box, $c > 0$ fixed$)$

$$
\begin{aligned}
-\Delta y + \mathcal{N}(\cdot, y) - u = 0 \ \text{ in } \Omega, &\quad \partial_\nu y = 0 \ \text{ on } \partial\Omega,\\
-\Delta p + \partial_y \mathcal{N}(\cdot, y) p + y = g \ \text{ in } \Omega, &\quad \partial_\nu p = 0 \ \text{ on } \partial\Omega,\\
-p + \lambda + \alpha u = 0 \ \text{ in } \Omega,&\\
\lambda - \max(0, \lambda + c(u - \overline{u})) - \min(0, \lambda + c(u - \underline{u})) = 0 \ \text{ in } \Omega,&
\end{aligned}
$$

- We use a semismooth Newton (SSN) method for solving the above system.
- The PDE is only fulfilled in the end of the iteration of the SSN.
- To respect the nature of the reduced problem, a SSN Sequential Quadratic Programming (SQP) algorithm is considered: For every k solve the (QP)

$$
\begin{aligned}
&\underset{\delta_u \in U}{\text{minimize}} \quad \langle \mathcal{J}'_{\mathcal{N}}(u_k) + \frac{1}{2} H_k(u_k)\delta_u, \delta_u \rangle_{U^*, U},\\
&\text{subject to } \ \underline{u} \le u_k + \delta_u \le \overline{u} \quad \text{a.e. in } \Omega.
\end{aligned}
$$

A SSN-SQP algorithm

Define a merit function $\Phi_k(\mu)$ as

$$\mathcal{J}_{\mathcal{N}}(u_k + \mu\delta_{u,k}) + \beta_k(\left\|(u_k + \mu\delta_{u,k} - \overline{u})^+\right\|_{L^2(\Omega)} + \left\|(u_k + \mu\delta_{u,k} - \underline{u})^-\right\|_{L^2(\Omega)}).$$

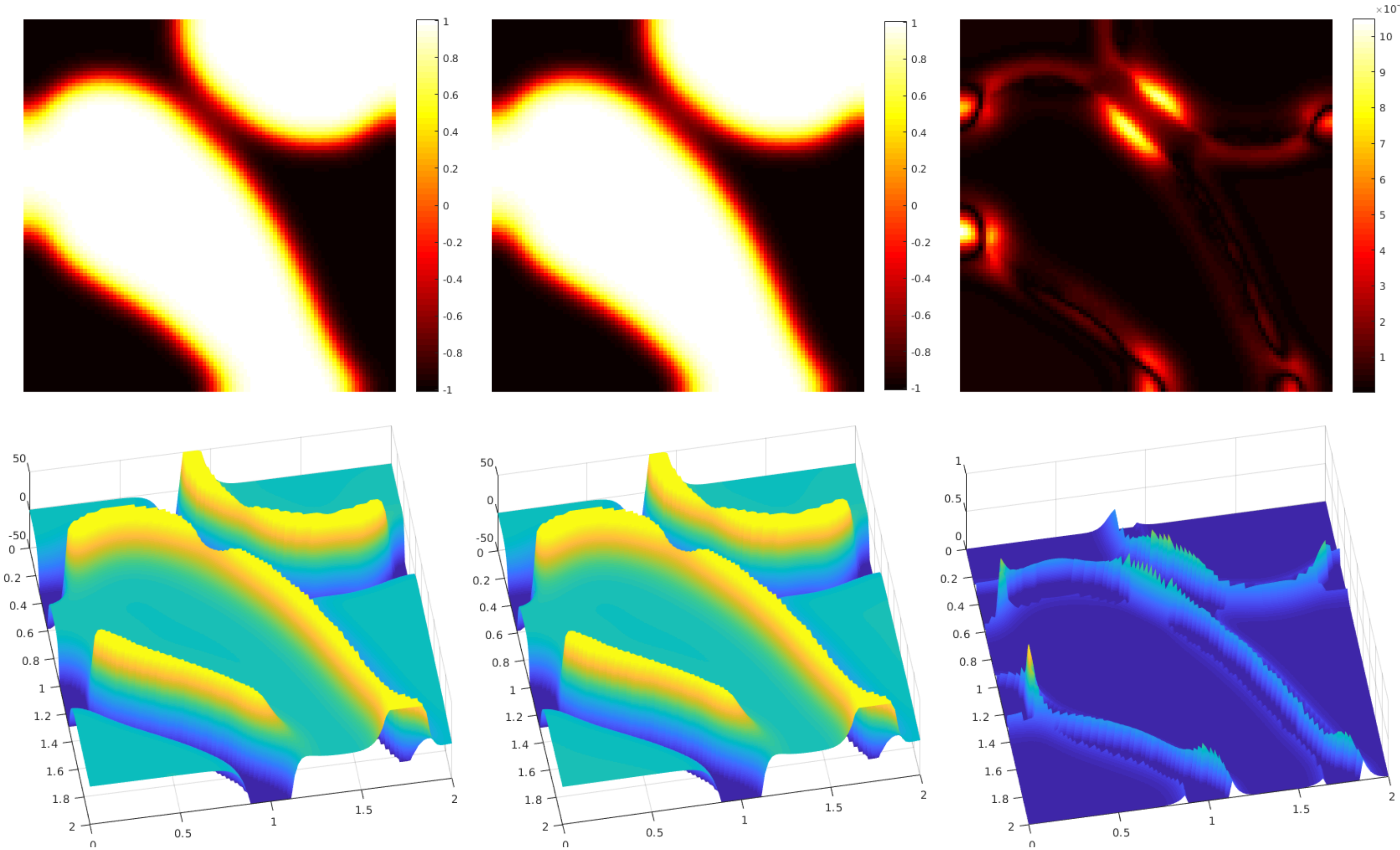
- Initialization: Using semi-smooth Newton for an initial guess of solutions.
- Key steps of every SQP:

(1) Compute an update direction $\delta_{u,k}$ using (inexact) SSN but to get approx. stat. point of QP.

(2) Using line search with Armijo condition to adjust step length $\mu_k > 0$ in every SQP sub-problem.
For every iteration l in the line search, to evaluate $\mathcal{J}_{\mathcal{N}}(u_k + \mu_k^l\delta_{u,k})$ we need the solution of the PDE which is obtained by Newton iterations.

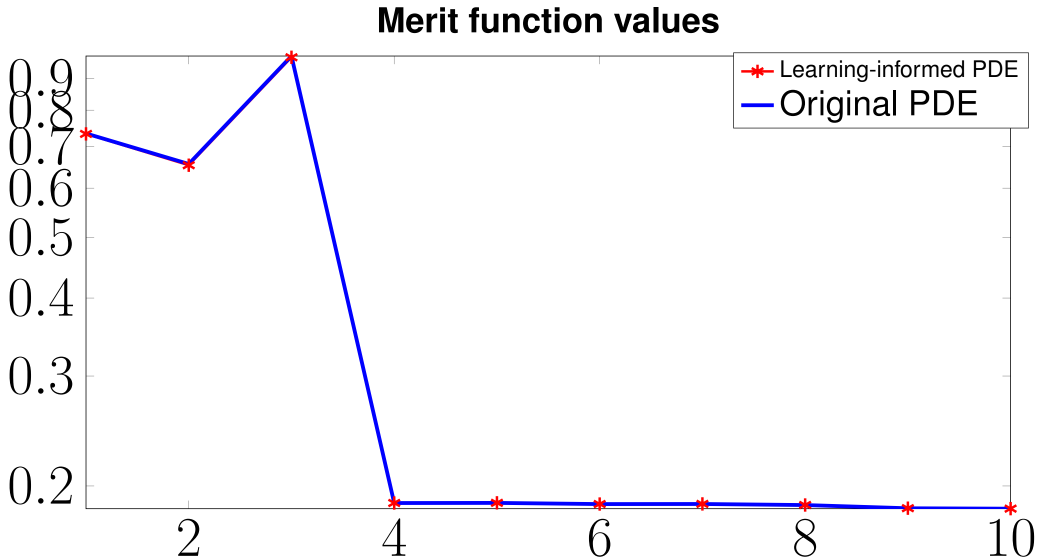
Primal-dual active set strategy (pdAS) is employed as SSN in every SQP sub-problem solve.

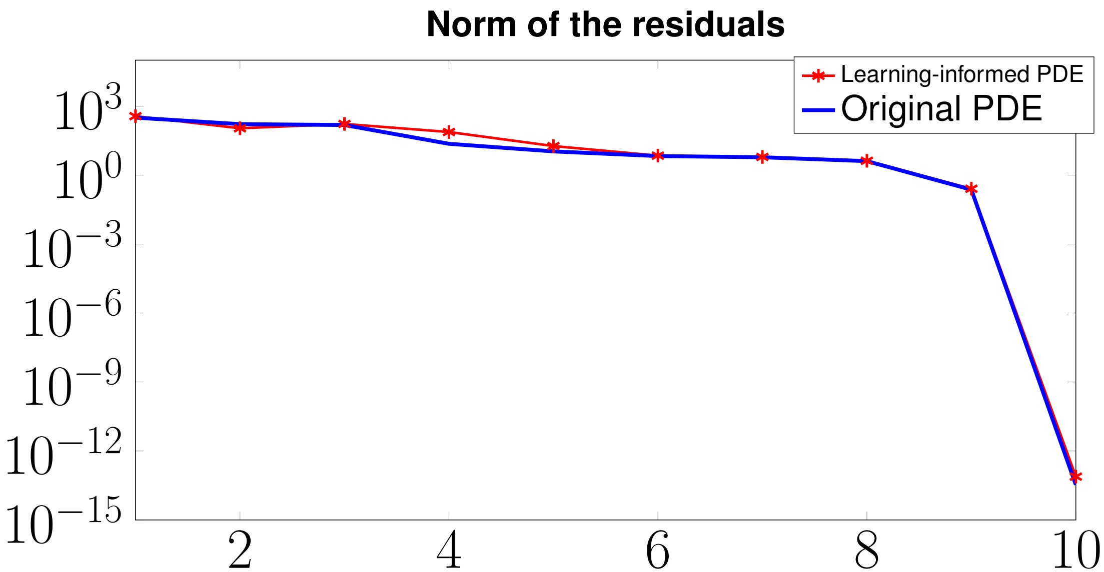
Example of stationary Allen-Cahn equation



Plots of state and control pairs $(y_{\mathcal{N}}, u_{\mathcal{N}})$ and (y^*, u^*) by learned (left) and exact (middle) PDEs, respectively, as well as their differences (right) $|y_{\mathcal{N}} - y^*|$, $|u_{\mathcal{N}} - u^*|$

Example of stationary Allen-Cahn equation





Optimal control of non-smooth PDEs

Consider now the following optimal control problem

$$\underset{(y,u)\in H^1(\Omega)\times L^2(\Omega)}{\text{minimize}} \quad \frac{1}{2}\|y-g\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)},$$

$$\text{subject to } -\Delta y + f(\cdot,y) = u \ \text{ in } \Omega, \quad y = 0 \ \text{ on } \partial\Omega.$$

$$u \in \mathcal{C}_{ad} := \{v \in L^2(\Omega) : a(x) \leq v(x) \leq b(x), \quad \text{for } x \in \Omega\}.$$

- The function $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is not necessarily Fréchet differentiable, but directional differentiable only.

[4] Christof, Meyer, Walther and Clason, Optimal control of a non-smooth semilinear elliptic equation, Mathematical Control & Related Fields, 2018

Optimal control of non-smooth PDEs

Consider now the following optimal control problem

$$\underset{(y,u)\in H^1(\Omega)\times L^2(\Omega)}{\text{minimize}} \quad \frac{1}{2}\|y-g\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)},$$

$$\text{subject to } -\Delta y + \mathcal{N}(\cdot, y) = u \quad \text{in } \Omega, \quad y = 0 \text{ on } \partial\Omega.$$

$$u \in \mathcal{C}_{ad} := \{v \in L^2(\Omega) : a(x) \le v(x) \le b(x), \quad \text{for } x \in \Omega\}.$$

- The function $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is not necessarily Fréchet differentiable, but directional differentiable only.
- f is learned via NNs with nonsmooth activation functions, e.g., ReLU, thus $\mathcal{N}$ is a nonsmooth function

---

Some relevant questions[4]:

- Various stationarity concepts and their relations
- Numerical algorithms for realizing the KKT condition (B-stationarity)

[4] Christof, Meyer, Walther and Clason, Optimal control of a non-smooth semilinear elliptic equation, Mathematical Control & Related Fields, 2018

Stationary conditions

- Primal optimality condition (B-stationarity condition): ($\Pi = \Pi_{\mathcal{N}}$)

$$(\Pi(\bar{u}) - g, \Pi'(\bar{u}; h)) + \alpha(\bar{u}, h) \geq 0 \quad \text{for all} \quad h \in \mathcal{T}_{\mathcal{C}_{\text{ad}}}(\bar{u})$$

  where

$$\mathcal{T}_{\mathcal{C}_{\text{ad}}}(\bar{u}) = \left\{ h \in L^2(\Omega) : h(x) \geq 0 \text{ a.e.} \bar{u}(x) = a(x), \quad h(x) \leq 0 \text{ a.e. } \bar{u}(x) = b(x) \right\}$$

- Dual optimality condition (C-stationarity condition):

$$\begin{aligned} -\Delta \bar{y} + \mathcal{N}(\cdot, \bar{y}) - \bar{u} = 0 \ \text{ in } \Omega, &\quad \bar{y} = 0 \ \text{ on } \partial\Omega, \\ -\Delta \bar{p} + \chi \bar{p} + \bar{y} = g \ \text{ in } \Omega, &\quad \bar{p} = 0 \ \text{ on } \partial\Omega, \\ \chi \in \partial_c \mathcal{N}(\cdot, \bar{y}) \text{ in } \Omega,& \\ (-\bar{p} + \alpha \bar{u}, u - \bar{u}) \geq 0 \ \text{ for all } \quad u \in \mathcal{C}_{ad}.& \end{aligned}$$

- Dual optimality condition (Strong-stationarity condition):
  C-stationarity + sign condition on the multiplier χ

$$\chi(x)\bar{p}(x) \in [\mathcal{N}'_+(x, y(x))\bar{p}(x), \mathcal{N}'_-(x, y(x))\bar{p}(x)] \quad \text{a.e. } x \in \Omega.$$

Relations among various stationary conditions

Let $\Omega_f \subset \Omega$ be the set where $f(\cdot, \bar{y})$ is **nondifferentiable**, and
$\Omega_{a,b} = \Omega_a \cup \Omega_b \subset \Omega$ be the **active set** where $\bar{u} = a$ or $\bar{u} = b$, and $a \leq b$ a.e. in Ω

At $(\bar{y}, \bar{u})$, the following constraint qualification is considered:
(i) Ω_f, Ω_a and Ω_b are measurable sets, resp.,
(ii) $|\Omega_f \cap \Omega_{a,b}| = 0$.

Selected results[5]:

- $(\bar{y}, \bar{u})$ locally optimal $\Rightarrow$ B-stationarity

[5]Master thesis: K. Völkner, supervisor: M. Hintermüller, Optimal control of a class of nonsmooth semilinear elliptic PDEs, 2021

Relations among various stationary conditions

Let $\Omega_f \subset \Omega$ be the set where $f(\cdot, \bar{y})$ is **nondifferentiable**, and $\Omega_{a,b} = \Omega_a \cup \Omega_b \subset \Omega$ be the **active set** where $\bar{u} = a$ or $\bar{u} = b$, and $a \leq b$ a.e. in Ω

At $(\bar{y}, \bar{u})$, the following constraint qualification is considered:
(i) Ω_f, Ω_a and Ω_b are measurable sets, resp.,
(ii) $|\Omega_f \cap \Omega_{a,b}| = 0$.

Selected results[5]:

- $(\bar{y}, \bar{u})$ locally optimal $\Rightarrow$ B-stationarity
- For piece-wise C^1 continuous function $f(x, \cdot)$, local optimality $\Rightarrow$ C-stationarity

[5]Master thesis: K. Völkner, supervisor: M. Hintermüller, Optimal control of a class of nonsmooth semilinear elliptic PDEs, 2021

Relations among various stationary conditions

Let $\Omega_f \subset \Omega$ be the set where $f(\cdot, \bar{y})$ is **nondifferentiable**, and $\Omega_{a,b} = \Omega_a \cup \Omega_b \subset \Omega$ be the **active set** where $\bar{u} = a$ or $\bar{u} = b$, and $a \leq b$ a.e. in Ω

At $(\bar{y}, \bar{u})$, the following constraint qualification is considered:
(i) Ω_f, Ω_a and Ω_b are measurable sets, resp.,
(ii) $|\Omega_f \cap \Omega_{a,b}| = 0$.

Selected results[5]:

- $(\bar{y}, \bar{u})$ locally optimal $\Rightarrow$ B-stationarity
- For piece-wise C^1 continuous function $f(x, \cdot)$, local optimality $\Rightarrow$ C-stationarity
- C-stationarity + constraint qualification $\Rightarrow$ strong stationarity

[5] Master thesis: K. Völkner, supervisor: M. Hintermüller, Optimal control of a class of nonsmooth semilinear elliptic PDEs, 2021

Relations among various stationary conditions

Let $\Omega_f \subset \Omega$ be the set where $f(\cdot, \bar{y})$ is **nondifferentiable**, and $\Omega_{a,b} = \Omega_a \cup \Omega_b \subset \Omega$ be the **active set** where $\bar{u} = a$ or $\bar{u} = b$, and $a \leq b$ a.e. in Ω

At $(\bar{y}, \bar{u})$, the following constraint qualification is considered:
(i) Ω_f, Ω_a and Ω_b are measurable sets, resp.,
(ii) $|\Omega_f \cap \Omega_{a,b}| = 0$.

Selected results[5]:

- $(\bar{y}, \bar{u})$ locally optimal $\Rightarrow$ B-stationarity
- For piece-wise C^1 continuous function $f(x, \cdot)$, local optimality $\Rightarrow$ C-stationarity
- C-stationarity + constraint qualification $\Rightarrow$ strong stationarity
- Strong-stationarity $\Rightarrow$ B-stationarity

[5]Master thesis: K. Völkner, supervisor: M. Hintermüller, Optimal control of a class of nonsmooth semilinear elliptic PDEs, 2021

Relations among various stationary conditions

Let $\Omega_f \subset \Omega$ be the set where $f(\cdot, \bar{y})$ is **nondifferentiable**, and
$\Omega_{a,b} = \Omega_a \cup \Omega_b \subset \Omega$ be the **active set** where $\bar{u} = a$ or $\bar{u} = b$, and $a \leq b$ a.e. in Ω

At $(\bar{y}, \bar{u})$, the following constraint qualification is considered:
(i) Ω_f, Ω_a and Ω_b are measurable sets, resp.,
(ii) $|\Omega_f \cap \Omega_{a,b}| = 0$.

Selected results[5]:

- $(\bar{y}, \bar{u})$ locally optimal $\Rightarrow$ B-stationarity
- For piece-wise C^1 continuous function $f(x, \cdot)$, local optimality $\Rightarrow$ C-stationarity
- C-stationarity + constraint qualification $\Rightarrow$ strong stationarity
- Strong-stationarity $\Rightarrow$ B-stationarity
- B-stationarity + constraint qualification $\Rightarrow$ strong-stationarity

For the last statement, the CQ requires that $\mathcal{T}_{\mathcal{C}_{\text{ad}}}(\bar{u})$ is dense in $L^2(\Omega)$.

[5] Master thesis: K. Völkner, supervisor: M. Hintermüller, Optimal control of a class of nonsmooth semilinear elliptic PDEs, 2021

Numerical algorithm– A descent method

Define an auxiliary problem[6]:

$$\min_h \quad \frac{1}{2}q(h,h) + (\Pi(u) - g, \Pi'(u,h)) + \alpha(u,h) \quad \text{over} \quad h \in \mathcal{F}. \tag{3.3}$$

Proposition

Let u be a feasible point for the reduced problem. Then the following properties are satisfied:

(1) *The problem* (3.3) *admits an optimal solution $\bar{h} \in \mathcal{T}_{\mathcal{C}_{ad}}(u)$.*

(2) *If $\bar{h} \neq 0$, then $\bar{h}$ is a descent direction for the reduced objective.*

(3) *If the directional derivative $\Pi'(u;\cdot) : L^p(\Omega) \to Y$ is bounded and linear, then $\bar{h}$ is unique.*

Conceptual algorithm: Solve Problem (3.3) iteratively using a line search method to find a descent direction of the reduced cost functional.

[6] Hintermüller, Surowiec. A bundle-free implicit programming approach for a class of elliptic MPECs in function space, Math. Prog. A, 2016.

Numerical algorithm– A descent method[7]

Consider a smooth approximation of the auxiliary problem:

$$\min_h \quad \frac{1}{2}q(h,h) + (\Pi(u) - g, \omega_\epsilon(u,h)) + \alpha(u,h) \quad \text{over} \quad h \in \mathcal{F}. \tag{3.4}$$

$\omega_\epsilon(u,h)$ takes into account the structure of the directional derivatives of ReLU network functions.

Lemma

Let u be a feasible point of the reduced problem. If $h \equiv 0$ solves (3.4) for all $\epsilon < \epsilon_0$, then u is a B-stationary point of the reduced problem.

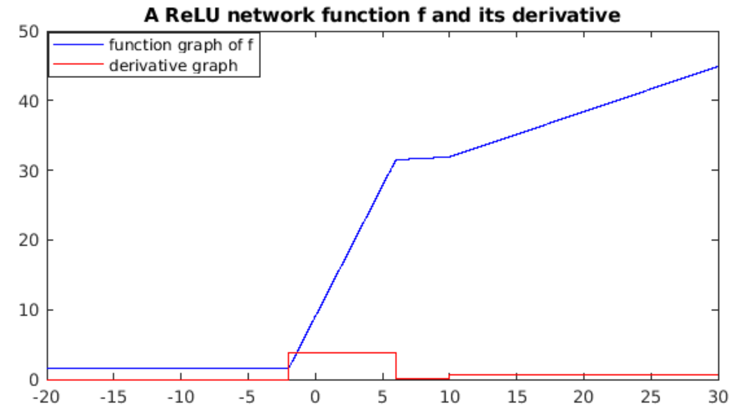
Proposition

Let u be a feasible point for the reduced optimal control problem. There exists $\epsilon^* > 0$, such that for all $\epsilon \leq \epsilon^*$, if $h_\epsilon \neq 0$ solves problem (3.4) at the feasible point u, then h_ϵ is a descent direction for the cost functional.

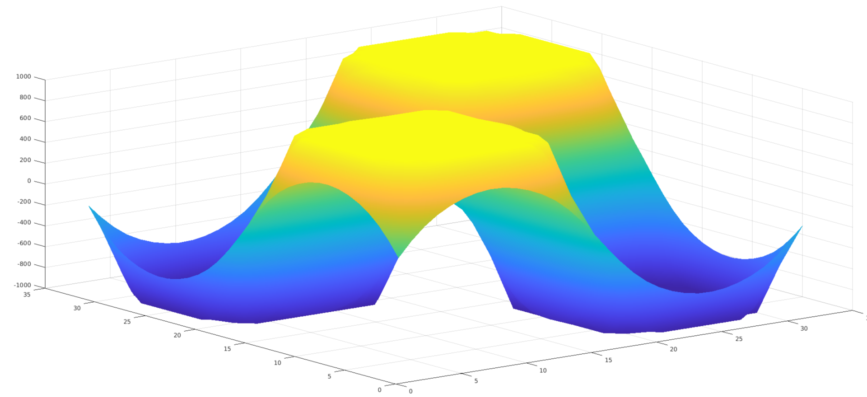
Algorithm: Primal-Dual-Active-Set algorithm + (semi-smooth) Newton method + Line search

[7] Dong, Hintermüller, Papafitsoros. Optimal control of learning-informed nonsmooth PDEs, 2021, in preparation.

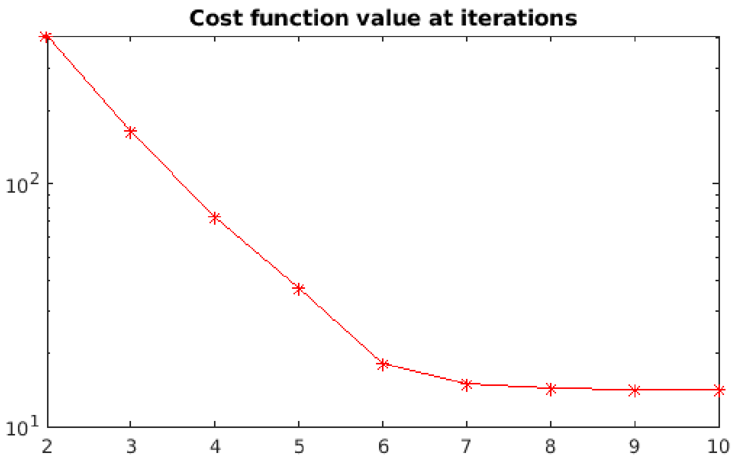
Numerical results



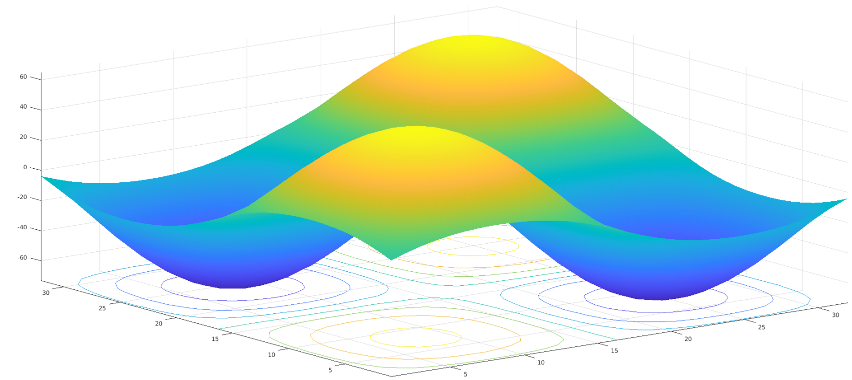
Monotone ReLU network function $\mathcal{N}$.



Computed control variable
($\alpha = 10^{-5}$, $a = -1000$, $b = 1000$)

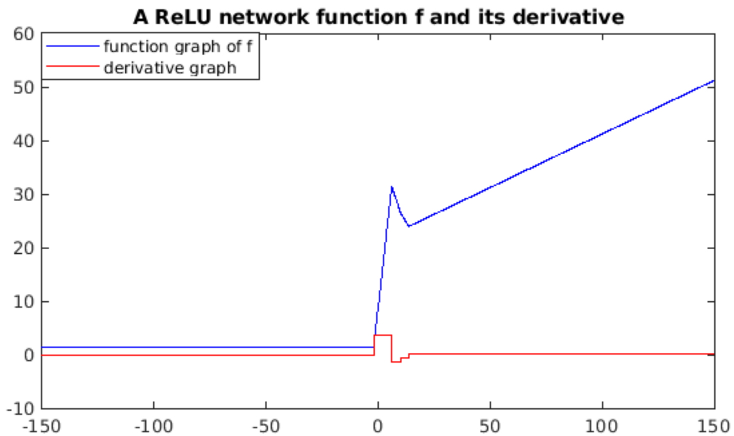


Cost function of the optimal control at iterations.

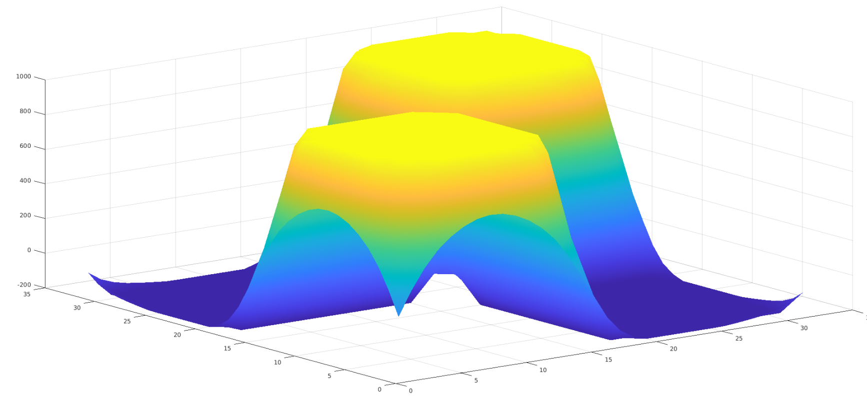


Computed state variable

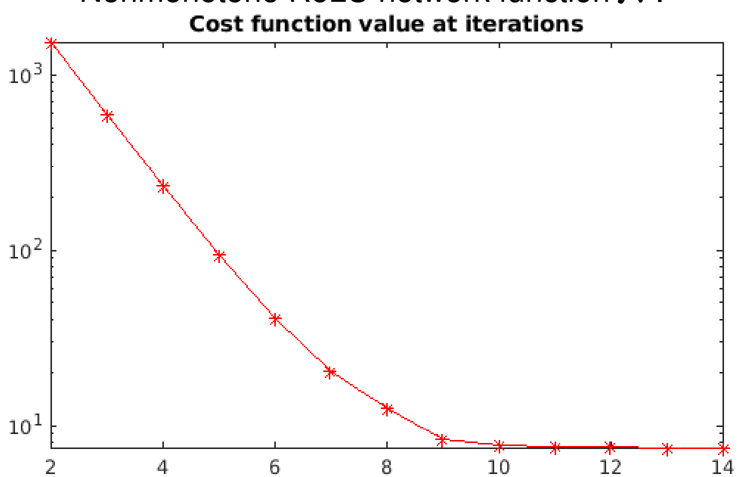
Numerical results



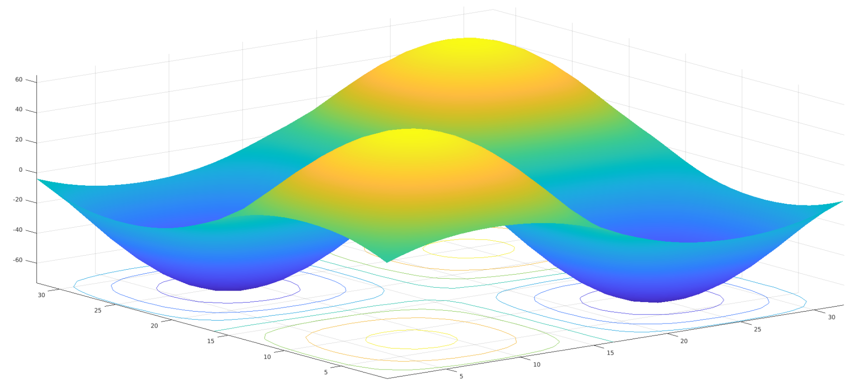
Nonmonotone ReLU network function $\mathcal{N}$.



Computed control variable
($\alpha = 10^{-5}$, $a = -200$, $b = 1000$).



Cost function of the optimal control at iterations.



Computed state variable.

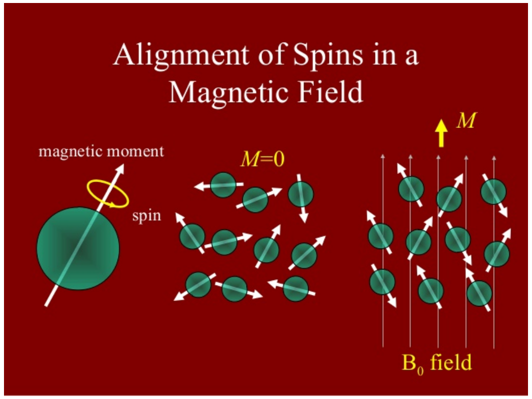
(Quantitative) MRI

Bloch equations describe the physical law behind MRI

$$\frac{\partial y}{\partial t}(t) = y(t) \times \gamma B(t) - \left(\frac{y_1(t)}{T_2}, \frac{y_2(t)}{T_2}, \frac{y_3(t) - \rho m_e}{T_1} \right),$$

where $B = B_0 + B_1 + G$ denotes magnetic field, ρ is proton density.
MRI experiment consists of three major steps:

- Aligning magnetic nuclear spins in an applied constant magnetic field B_0
- Perturbing this alignment via radio frequency (RF) pulse B_1
- Applying magnetic gradient field G to distinguish individual contributions



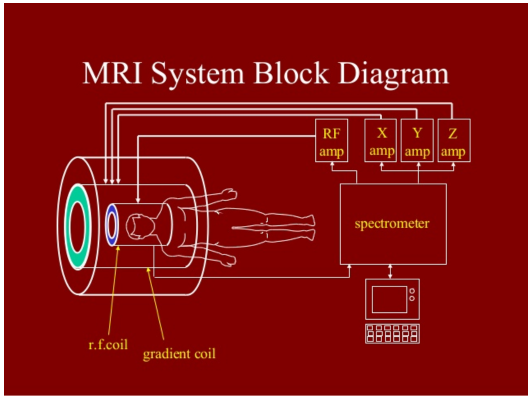


Figure: MRI diagram (Published in Health and Medicine)

Bloch equations describe the physical law behind MRI

$$\frac{\partial y}{\partial t}(t) = y(t) \times \gamma B(t) - \left(\frac{y_1(t)}{T_2}, \frac{y_2(t)}{T_2}, \frac{y_3(t) - \rho m_e}{T_1} \right),$$

where $B = B_0 + B_1 + G$ denotes magnetic field, ρ is proton density.

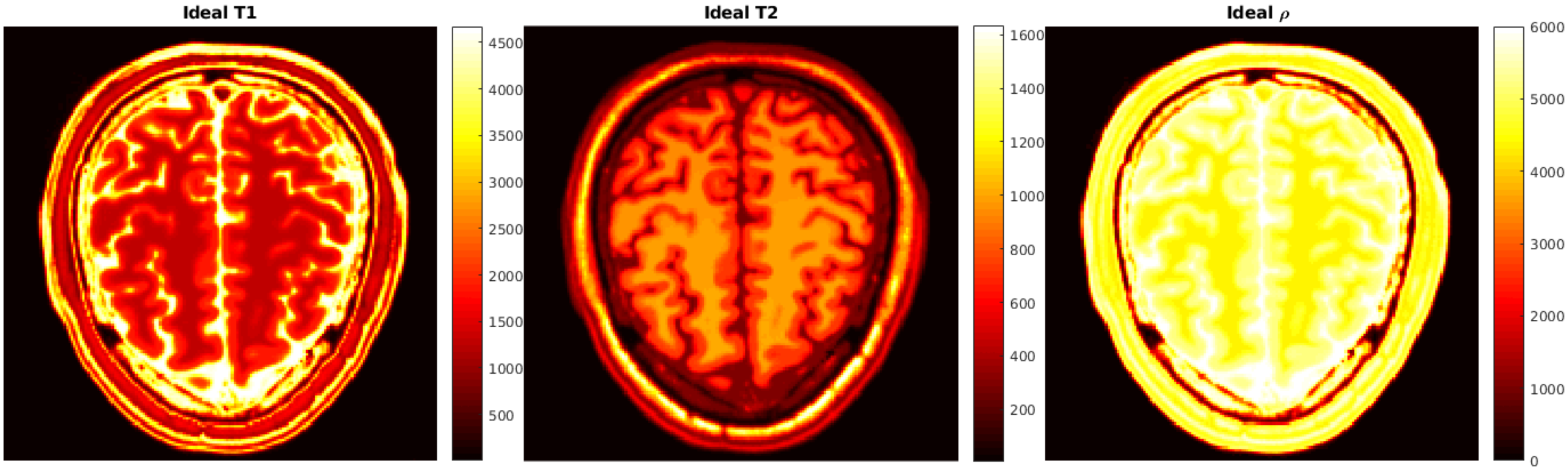


Figure: Simulated ideal tissue parameters of a brain phantom.

qMRI as a „control problem”

qMRI fits the general framework:

$$\underset{(y,u)}{\text{minimize}} \quad \frac{1}{2}\|P\mathcal{F}(y) - g^\delta\|_H^2 + \frac{\alpha}{2}\|u\|_U^2,$$

subject to

$$\frac{\partial y}{\partial t}(t) = y(t) \times \gamma B(t) - \left(\frac{y_1(t)}{T_2}, \frac{y_2(t)}{T_2}, \frac{y_3(t) - \rho m_e}{T_1}\right), \; t = t_1, \ldots, t_L,$$
$$y(0) = \rho m_0,$$
$$u \in \mathcal{C}_{ad}.$$

- The goal is to estimate the physical parameters $u = (\rho, T_1, T_2)$

qMRI as a „control problem”

qMRI fits the general framework:

$$\underset{(y,u)}{\text{minimize}} \quad \frac{1}{2}\|P\mathcal{F}(y) - g^\delta\|_H^2 + \frac{\alpha}{2}\|u\|_U^2,$$

subject to

$$y = \mathcal{N}(u),$$
$$u \in \mathcal{C}_{ad}.$$

- The goal is to estimate the physical parameters $u = (\rho, T_1, T_2)$
- ANNs $\mathcal{N}$ approximate the parameter-to-solution map (Nemytskii type):

$$(\rho, T_1, T_2) \mapsto (y_{t_1}, \ldots, y_{t_L})$$

Universal approximation of learning-informed Bloch operator

Proposition

The operator $\Pi : \mathcal{C}_{ad} \subset [L^\infty_\epsilon(\Omega)^+]^3 \to [(L^\infty(\Omega))^3]^L$ *is Lipschitz continuous, and Fréchet differentiable with locally Lipschitz derivative.*

Both Π and $\Pi_\mathcal{N} = \mathcal{N}$ are operators of Nemytskii type in the qMRI case.

Proposition

Let $u = (T_1, T_2, \rho)^\top \in \mathcal{C}_{ad}$*. Then for arbitrary small* $\epsilon > 0$ *and* $\epsilon_1 > 0$*, there always exist neural network approximations so that*

$$\|\Pi_\mathcal{N}(u) - \Pi(u)\|_{[L^\infty(\Omega)^3]^L} \leq \epsilon,$$

and

$$\|\Pi'_\mathcal{N}(u) - \Pi'(u)\|_{\mathcal{L}([L^2(\Omega)]^3,[L^\infty(\Omega)^3]^L)} \leq \epsilon_1,$$

are satisfied.

SQP algorithm

Define

$$\mathcal{J}_{\mathcal{N}}(u) := \frac{1}{2}\|P\mathcal{F}(\mathcal{N}(u)) - g^\delta\|_H^2 + \frac{\alpha}{2}\|u\|_U^2.$$

The derivative $\mathcal{J}'_{\mathcal{N}}(u)$ has an explicit form

$$(\rho(\mathcal{N}'(T_1,T_2))^*, \mathcal{N}(T_1,T_2))^\top \mathcal{F}^*(\mathcal{F}(\rho\mathcal{N}(T_1,T_2)) - g) + \alpha(\mathsf{Id} - \Delta)(T_1,T_2,\rho)^\top.$$

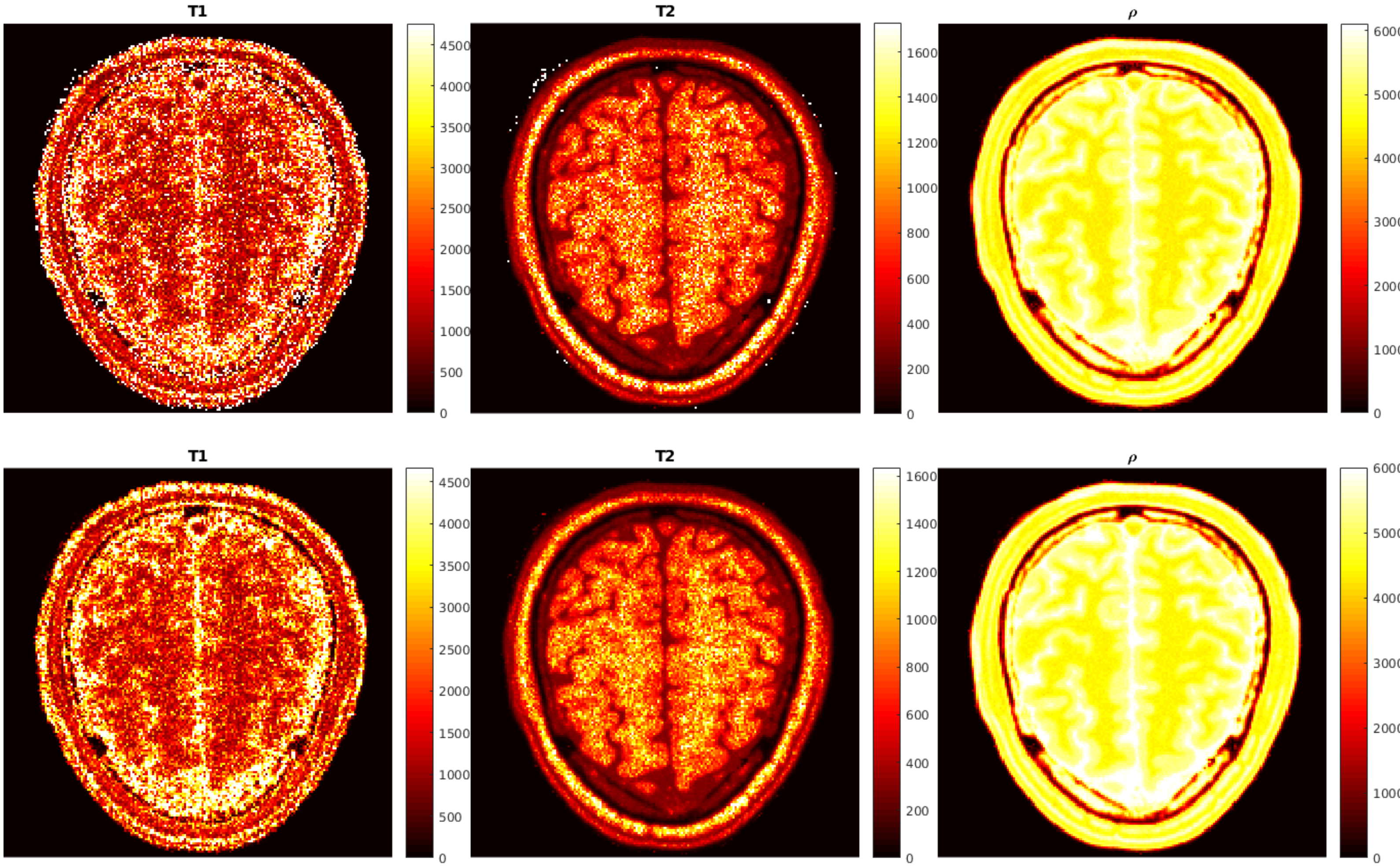
Every QP-step solves

$$\begin{aligned} &\text{minimize} \quad \langle \mathcal{J}'_{\mathcal{N}}(u_k), h\rangle_{U^*,U} + \frac{1}{2}\langle H_k(u_k)h, h\rangle_{U^*,U} \quad \text{over } h \in U \\ &\text{s.t.} \quad u_k + h \in \mathcal{C}_{ad}, \end{aligned}$$

where $H_k(u_k)$ is a pos.-def. approx. of the Hessian of $\mathcal{J}_{\mathcal{N}}$ at $u_k \in \mathcal{C}_{ad}$:

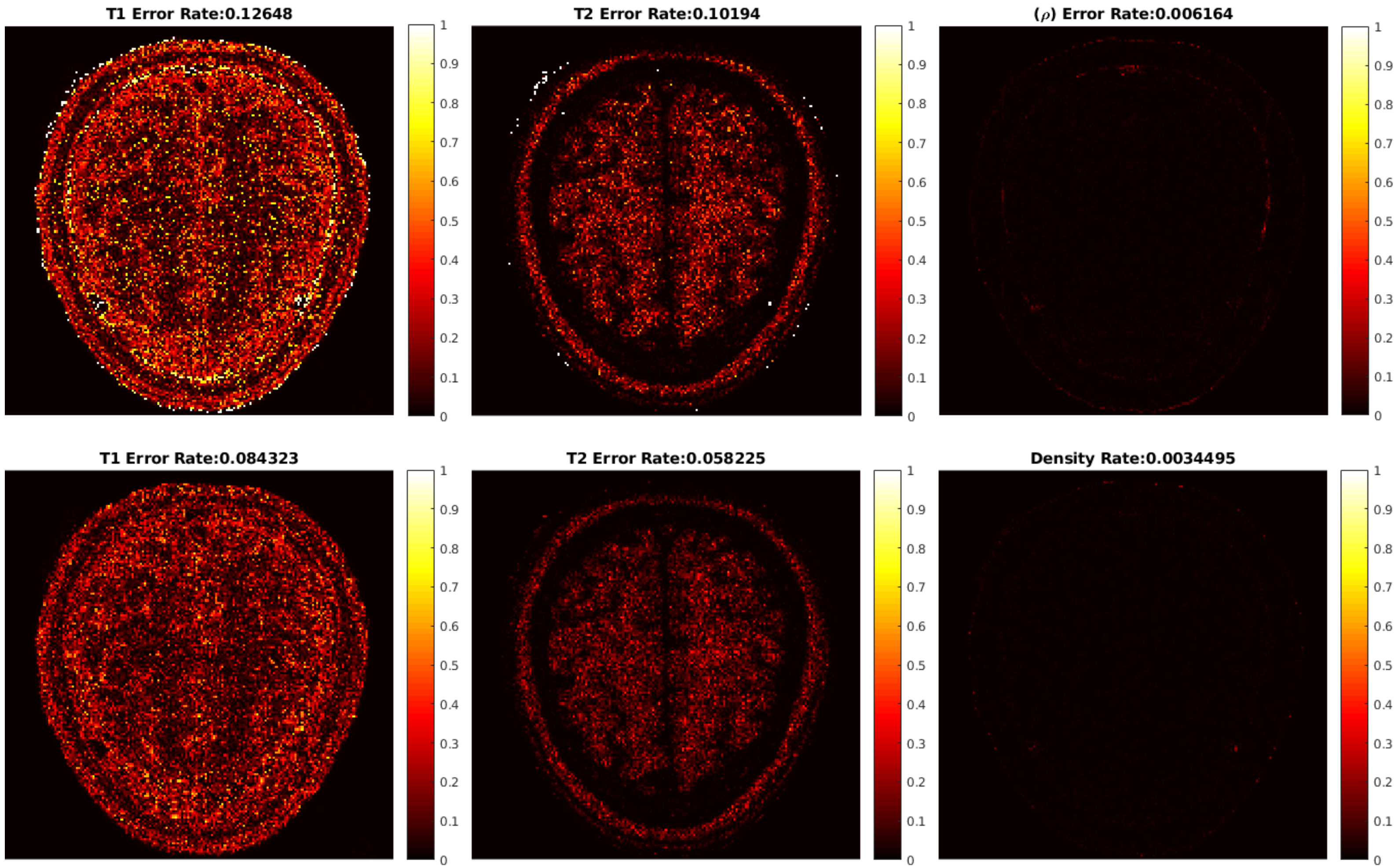
$$(\rho(\mathcal{N}'(T_1,T_2))^*, \mathcal{N}(T_1,T_2))^\top \mathcal{F}^*\mathcal{F}(\rho(\mathcal{N}'(T_1,T_2)), \mathcal{N}(T_1,T_2)) + \alpha(\mathsf{Id} - \Delta).$$

Quantitative magnetic resonance imaging



Learning-based (bottom) compared to ab initio physics-integrated method (above)

Quantitative magnetic resonance imaging



Learning-based (bottom) compared to a pure physics-integrated method (above)

Conclusion

What we offer:

- A generic optimization framework with learning-informed physical constraints
- Both analysis and numerical algorithms for the overall optimization framework
- Learning specific operators between infinite dimensional spaces
- Universal approximation properties for the learning-informed operators
- The framework for learning-informed *nonsmooth* physical models

Ongoing:

- More general physical operator learning schemes
- Interplay of operator learning and optimal control
- Hybrid physics-informed NN for multi-scale problems

Thank you!