Stochastic Algorithms for Constrained Optimization for Informed Learning

Frank E. Curtis, Lehigh University

presented at

INFORMS Annual Meeting

Seattle, Washington

October 20, 2024

Outline

Motivation

Informed Learning

Discussion

Conclusion

Outline

Motivation

Informed Learning

Discussion

Conclusion

Constrained continuous optimization

Consider the setting of solving constrained continuous optimization problems of the form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \leq 0 \end{aligned}$$

when at any $x \in \mathbb{R}^n$ one has that

- $c_{\mathcal{E}}(x)$ and $c_{\mathcal{I}}(x)$ can be computed exactly
- $\nabla c_{\mathcal{E}}(x)$ and $\nabla c_{\mathcal{I}}(x)$ can be computed exactly
- $f(x)$ and $\nabla f(x)$ cannot be computed exactly—only have (unbiased) estimates

References

- A. S. Berahas, F. E. Curtis, D. P. Robinson, and B. Zhou, "Sequential Quadratic Optimization for Nonlinear Equality Constrained Stochastic Optimization," *SIAM Journal on Optimization*, 31(2):1352–1379, 2021.
- A. S. Berahas, F. E. Curtis, M. J. O'Neill, and D. P. Robinson, "A Stochastic Sequential Quadratic Optimization Algorithm for Nonlinear Equality Constrained Optimization with Rank-Deficient Jacobians," *Mathematics of Operations Research*, https://doi.org/10.1287/moor.2021.0154, 2023.
- F. E. Curtis, D. P. Robinson, and B. Zhou, "Inexact Sequential Quadratic Optimization for Minimizing a Stochastic Objective Subject to Deterministic Nonlinear Equality Constraints," to appear in *INFORMS Journal on Optimization*, https://arxiv.org/abs/2107.03512.
- F. E. Curtis, M. J. O'Neill, and D. P. Robinson, "Worst-Case Complexity of an SQP Method for Nonlinear Equality Constrained Stochastic Optimization," *Mathematical Programming*, https://doi.org/10.1007/s10107-023-01981-1, 2023.
- F. E. Curtis, S. Liu, and D. P. Robinson, "Fair Machine Learning through Constrained Stochastic Optimization and an ϵ-Constraint Method," *Optimization Letters*, https://doi.org/10.1007/s11590-023-02024-6, 2023.
- F. E. Curtis, X. Jiang, and Q. Wang, "Almost-sure convergence of iterates and multipliers in stochastic sequential quadratic optimization," https://arxiv.org/abs/2308.03687.
- F. E. Curtis, D. P. Robinson, and B. Zhou, "Sequential Quadratic Optimization for Stochastic Optimization with Deterministic Nonlinear Inequality and Equality Constraints," https://arxiv.org/abs/2302.14790.
- F. E. Curtis, V. Kungurtsev, D. P. Robinson, and Q. Wang, "A Stochastic-Gradient-based Interior-Point Algorithm for Solving Smooth Bound-Constrained Optimization Problems," https://arxiv.org/abs/2304.14907.

Where do we go from here?

There are many open questions that are of interest to optimizers such as

- other algorithm variants with same guarantees
- strengthened guarantees (e.g., other growth conditions, convex settings)
- improved worst-case complexity properties
- loosened constraint qualification requirements
- second-order-type methods
- generalization properties
- trade-off analyses (Bottou–Bosquet)

Outline

Motivation

Informed Learning

Discussion

Conclusion

Learning: Prediction function

Aim: Determine a prediction function p from a family $\mathcal{P}$ such that

$$p(a_j)$$

yields an accurate prediction corresponding to any given input feature vector a_j.

Learning: Prediction function, parameterized

Let us say that the family is parameterized by some vector x such that

$$p(a_j, x)$$

yields an accurate prediction corresponding to any given input feature vector a_j.

Learning: Supervised

In *supervised* learning, we have known input-output pairs $\{(a_j, b_j)\}_{j=1}^{n_o}$. Then,

$$\min_{x \in \mathbb{R}^n} \frac{1}{n_o} \sum_{j=1}^{n_o} \ell(p(a_j, x), b_j)$$

becomes our empirical-loss training problem to determine the *optimal* x.

Learning: Supervised and regularized

If we aim to impose some structure on the solution x, then we may consider

$$\min_{x \in \mathbb{R}^n} \; \frac{1}{n_o} \sum_{j=1}^{n_o} \ell(p(a_j, x), b_j) + r(x)$$

where r is a *regularization* function.

Learning: Supervised and regularized

If we aim to impose some structure on the solution x, then we may consider

$$\min_{x \in \mathbb{R}^n} \frac{1}{n_o} \sum_{j=1}^{n_o} \ell(p(a_j, x), b_j) + r(x)$$

where r is a *regularization* function. Is this good for *informed* learning?

Learning: Supervised and informed through model design

One approach is to embed information in the prediction function itself, so

$$\min_{x \in \mathbb{R}^n} \; \frac{1}{n_o} \sum_{j=1}^{n_o} \ell(p(a_j, x), b_j)$$

ensures that information is enforced with every forward pass. (Is this enough and/or efficient?)

Learning: Supervised and informed with *soft* constraints

Added to the loss (e.g., mean-squared error), we might consider

$$\min_{x\in\mathbb{R}^n} \; \frac{1}{n_o}\sum_{j=1}^{n_o} \ell(p(a_j, x), b_j) + \frac{1}{n_c}\sum_{j=1}^{n_c} \phi(p(\tilde{a}_j, x), \dots, \tilde{b}_j)$$

where $\{(\tilde{a}_j, \tilde{b}_j)\}_{j=1}^{n_c}$ are known input-output pairs and ϕ encodes information.

Learning: Supervised and informed with *hard* constraints

Alternatively (or in addition), how about *hard* constraints during training, as in

$$\min_{x\in\mathbb{R}^n} \ \frac{1}{n_o}\sum_{j=1}^{n_o}\ell(p(a_j,x),b_j)+\frac{1}{n_c}\sum_{j=1}^{n_c}\phi(p(\tilde{a}_j,x),\dots,\tilde{b}_j)$$
$$\text{s.t.} \ \ \varphi(p(\tilde{a}_j,x),\dots,\tilde{b}_j)=0 \ (\text{or} \leq 0) \ \ \text{for all} \ \ i\in\{1,\dots,n_c\}$$

such that we restrict attention to functions that are informed implicitly?

Expected-loss training problems

For the sake of generality/generalizability, the expected-loss objective function is

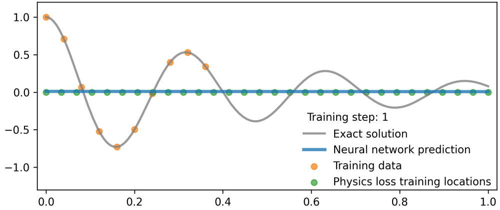
$$\int_{\mathcal{A}\times\mathcal{B}} \ell(p(a,x),b)\mathrm{d}\mathbb{P}(a,b) \equiv \mathbb{E}_\omega[F(x,\omega)] =: f(x)$$

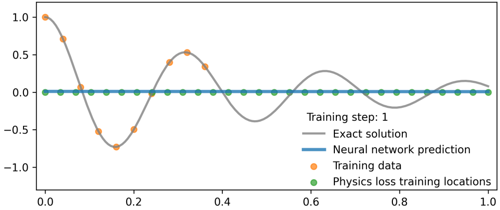
Assuming values and derivatives can be computed, the constraints are

$$c_\mathcal{E}(x) = 0 \quad \text{and} \quad c_\mathcal{I}(x) \leq 0$$

e.g., imposing a fixed set of constraints corresponding to a fixed set of sample data

Predicting movement of a spring





Problem from https://benmoseley.blog/blog/

Outline

Motivation

Informed Learning

Discussion

Conclusion

Topic #1: Data-driven constraints

The aforementioned approaches struggle with many data-driven constraints.

What other problem formulations should be considered?

- progressively more constraints
- expectation constraints
- probabilistic constraints
- noisy constraints

Various modeling issues arise:

- $c_i(x) = 0$ for all $i = 0, 1, 2, \ldots$ (infeasible?)
- $\frac{1}{N} \sum_{i \in [N]} c_i(x) = 0$ (too weak?)
- $c_i(x) = 0$ for at least $M \in [N]$ indices in $[N]$ (combinatorial issues? hard to choose M?)
- $c_i(x) = 0$ for all $i \in [N]$ ideally, but satisfied with $|c_i(x)| \leq \epsilon$ for all $i \in [N]$
- …different from $|c_i(x)| \leq \epsilon$ as inequality constraints

Topic #2: Algorithms that might actually be useful

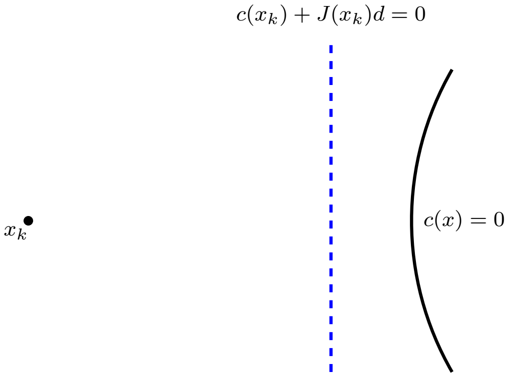
One should not consider the problem formulation in isolation.

What algorithms should be considered?

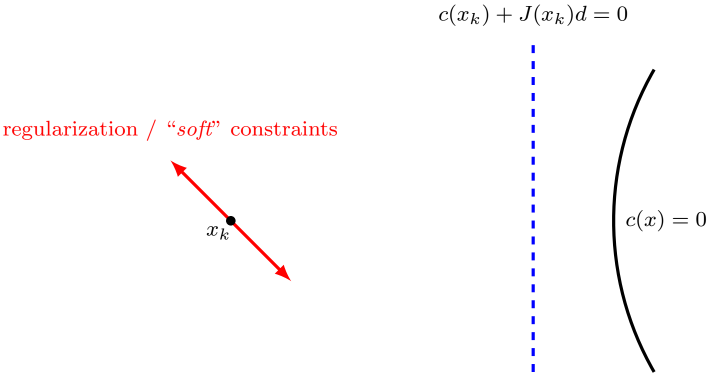
- feasible methods (impractical)
- alternating methods (no evidence of good practical performance)
- penalty and augmented Lagrangian methods (might as well not call them constraints)
- Newton-based methods for constraints ... ???

$$\begin{bmatrix} \mathrm{diag}(\cdot) & J_k^T \\ J_k & 0 \end{bmatrix}$$

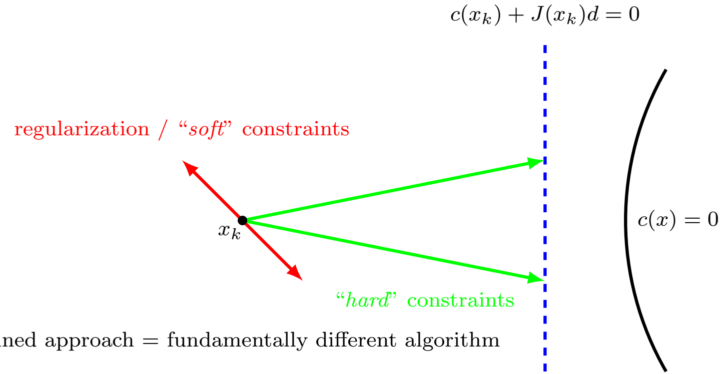
SQP illustration



SQP illustration



SQP illustration

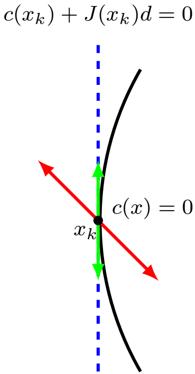


constrained approach = fundamentally different algorithm

SQP illustration

regularization / "*soft*" constraints

"*hard*" constraints $\Longrightarrow$ step in null space



Outline

Motivation

Informed Learning

Discussion

Conclusion

Summary

A good-sized body of work on stochastic-gradient-based methods for constrained optimization.

- practical methods
- convergence and complexity guarantees
- ... numerous open questions remain

However, we should think beyond the “starting point” formulation.

- data-driven constraint formulations
- corresponding algorithms that may be useful in practice

References

- A. S. Berahas, F. E. Curtis, D. P. Robinson, and B. Zhou, "Sequential Quadratic Optimization for Nonlinear Equality Constrained Stochastic Optimization," *SIAM Journal on Optimization*, 31(2):1352–1379, 2021.
- A. S. Berahas, F. E. Curtis, M. J. O'Neill, and D. P. Robinson, "A Stochastic Sequential Quadratic Optimization Algorithm for Nonlinear Equality Constrained Optimization with Rank-Deficient Jacobians," *Mathematics of Operations Research*, https://doi.org/10.1287/moor.2021.0154, 2023.
- F. E. Curtis, D. P. Robinson, and B. Zhou, "Inexact Sequential Quadratic Optimization for Minimizing a Stochastic Objective Subject to Deterministic Nonlinear Equality Constraints," to appear in *INFORMS Journal on Optimization*, https://arxiv.org/abs/2107.03512.
- F. E. Curtis, M. J. O'Neill, and D. P. Robinson, "Worst-Case Complexity of an SQP Method for Nonlinear Equality Constrained Stochastic Optimization," *Mathematical Programming*, https://doi.org/10.1007/s10107-023-01981-1, 2023.
- F. E. Curtis, S. Liu, and D. P. Robinson, "Fair Machine Learning through Constrained Stochastic Optimization and an ϵ-Constraint Method," *Optimization Letters*, https://doi.org/10.1007/s11590-023-02024-6, 2023.
- F. E. Curtis, D. P. Robinson, and B. Zhou, "Sequential Quadratic Optimization for Stochastic Optimization with Deterministic Nonlinear Inequality and Equality Constraints," https://arxiv.org/abs/2302.14790.
- F. E. Curtis, V. Kungurtsev, D. P. Robinson, and Q. Wang, "A Stochastic-Gradient-based Interior-Point Algorithm for Solving Smooth Bound-Constrained Optimization Problems," https://arxiv.org/abs/2304.14907.

Thank you!

Questions?