SIAG/OPT Views and News

A Forum for the SIAM Activity Group on Optimization

Volume 26 Number 1 May 2018

Contents

Article

This issue's article highlights work presented at a Workshop on Continuous Optimization at the 2017 Foundations of Computational Mathematics conference held in Barcelona, Spain from July 17th to 19th, 2017.

The Proximal Point Method Revisited

Dmitriy Drusvyatskiy
University of Washington
Department of Mathematics
Seattle, WA 98195, USA
ddrusv@uw.edu
www.math.washington.edu/~ddrusv/

1 Introduction

The proximal point method is a conceptually simple algorithm for minimizing a function f on $\mathbb{R}^d$. Given an iterate x_t, the method defines x_{t+1} to be any minimizer of the proximal subproblem

$$\operatorname*{argmin}_x \ \left\{ f(x) + \tfrac{1}{2\nu}\|x - x_t\|^2 \right\},$$

for an appropriately chosen parameter $\nu > 0$. At first glance, each proximal subproblem seems no easier than minimizing f in the first place. On the contrary, the addition of the quadratic penalty term often regularizes the proximal subproblems and makes them well conditioned. Case in point, the subproblem may become convex despite f not being convex; and even if f were convex, the subproblem has a larger strong convexity parameter, thereby facilitating faster numerical methods.

Despite the improved conditioning, each proximal subproblem still requires invoking an iterative solver. For this reason, the proximal point method has predominantly been thought of as a theoretical or conceptual algorithm, only guiding algorithm design and analysis rather than being implemented directly. One good example is the proximal bundle method [43], which approximates each proximal subproblem by a cutting-plane model. In the past few years, this viewpoint has undergone a major revision. In a variety of circumstances, the proximal point method (or a close variant) with a judicious choice of the control parameter $\nu > 0$ and an appropriate iterative method for the subproblems can lead to practical and theoretically sound numerical methods.

In this article, I will briefly describe three recent examples of this trend:

- a subgradient method for weakly convex expectation minimization problems [24],
- the prox-linear algorithm for minimizing compositions of convex functions and smooth maps [11,16,29,31,44,54], and
- the Catalyst generic acceleration schema [45] for regularized empirical risk minimization.

I will focus only on the proximal point method for minimizing functions, as outlined above. The proximal point methodology applies much more broadly to monotone operator inclusions; I refer the reader to the monograph of Bauschke and Combette [7] or the seminal work of Rockafellar [61].

2 Notation

The following two constructions will play a basic role in the article. For any closed function f on $\mathbb{R}^d$, the *Moreau envelope* and the *proximal map* are

$$f_\nu(z) := \inf_x \left\{ f(x) + \tfrac{1}{2\nu}\|x-z\|^2 \right\},$$
$$\mathrm{prox}_{\nu f}(z) := \operatorname*{argmin}_x \left\{ f(x) + \tfrac{1}{2\nu}\|x-z\|^2 \right\},$$

respectively. In this notation, the proximal point method is simply the fixed-point recurrence on the proximal map:[1]

$$\textbf{Step}\, t: \qquad \text{choose } x_{t+1} \in \mathrm{prox}_{\nu f}(x_t).$$

Clearly, to have any hope of solving the proximal subproblems, one must ensure that they are convex. Consequently, the class of weakly convex functions forms the natural setting for the proximal point method.

Definition 1. A function f is called *ρ-weakly convex* if the assignment $x \mapsto f(x) + \frac{\rho}{2}\|x\|^2$ is a convex function.

For example, a C^1-smooth function with ρ-Lipschitz gradient is ρ-weakly convex, while a C^2-smooth function f is ρ-weakly convex precisely when the minimal eigenvalue of its Hessian is uniformly bounded below by $-\rho$. In essence, weak convexity precludes functions that have downward kinks. For instance, $f(x) := -\|x\|$ is not weakly convex since no addition of a quadratic makes the resulting function convex.

Whenever f is ρ-weakly convex and the proximal parameter ν satisfies $\nu < \rho^{-1}$, each proximal subproblem is itself convex and therefore globally tractable. Moreover, in this setting, the Moreau envelope is C^1-smooth with gradient

$$\nabla f_\nu(x) = \nu^{-1}(x - \mathrm{prox}_{\nu f}(x)). \tag{1}$$

Rearranging the gradient formula yields the useful interpretation of the proximal point method as gradient descent on the Moreau envelope

$$x_{t+1} = x_t - \nu \nabla f_\nu(x_t).$$

[1] In order to ensure that $\mathrm{prox}_{\nu f}(\cdot)$ is nonempty, it suffices to assume that f is bounded from below.

In summary, the Moreau envelope f_ν serves as a C^1-smooth approximation of f for all small ν. Moreover, the two conditions

$$\|\nabla f_\nu(x_t)\| < \varepsilon$$

and

$$\|\nu^{-1}(x_t - x_{t+1})\| < \varepsilon,$$

are equivalent for the proximal point sequence $\{x_t\}$. Hence, the step size $\|x_t - x_{t+1}\|$ of the proximal point method serves as a convenient termination criterion.

Examples of weakly convex functions

Weakly convex functions are widespread in applications and are typically easy to recognize. One common source of weakly convex functions is the composite problem class:

$$\min_x \; F(x) := g(x) + h(c(x)), \tag{2}$$

where $g\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a closed convex function, $h\colon \mathbb{R}^m \to \mathbb{R}$ is convex and L-Lipschitz, and $c\colon \mathbb{R}^d \to \mathbb{R}^m$ is a C^1-smooth map with β-Lipschitz gradient. An easy argument shows that the composite function F is $L\beta$-weakly convex. This is a worst-case estimate. In concrete circumstances, the composite function F may have a much more favorable weak convexity constant. This is the case for instance in phase retrieval [32, Section 3.2]; see Example 2.4 for the problem definition.

Example 2.1 (Additive composite)**.** The most prevalent example is additive composite minimization. In this case, the map c maps to the real line, and h is the identity function:

$$\min_x \; c(x) + g(x). \tag{3}$$

Such problems appear often in statistical learning and imaging. Various specialized algorithms are available; see, for example, [8] or [55].

Example 2.2 (Nonlinear least squares)**.** The composite problem class also captures nonlinear least-squares problems with bound constraints:

$$\min_x \; \|c(x)\|_2 \quad \text{subject to} \quad l_i \le x_i \le u_i \;\forall i.$$

Such problems pervade engineering and scientific applications.

Example 2.3 (Exact penalty formulations)**.** Consider a nonlinear optimization problem:

$$\min_x \; \{f(x) : G(x) \in \mathcal{K}\},$$

where f and G are smooth maps and $\mathcal{K}$ is a closed convex cone. An accompanying *penalty formulation*—ubiquitous in nonlinear optimization—takes the form

$$\min_x \; f(x) + \lambda \cdot \mathrm{dist}_{\mathcal{K}}(G(x)),$$

where $\mathrm{dist}_{\mathcal{K}}(\cdot)$ is the distance to $\mathcal{K}$ in some norm. Historically, exact penalty formulations served as the early motivation for the class (2).

Example 2.4 (Robust phase retrieval)**.** Phase retrieval is a common computational problem, with applications in diverse areas such as imaging, X-ray crystallography, and speech processing. For simplicity, I will focus on the version of the problem over the reals. The (real-valued) phase retrieval problem seeks to determine a point x satisfying the magnitude conditions,

$$|\langle a_i, x\rangle| \approx b_i \quad \text{for } i = 1, \ldots, m,$$

where $a_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$ are given. Whenever gross outliers occur in the measurements b_i, the following robust formulation of the problem is appealing [23, 32, 34]:

$$\min_x \ \frac{1}{m}\sum_{i=1}^m |\langle a_i, x\rangle^2 - b_i^2|.$$

Clearly, this is an instance of (2). For some recent perspectives on phase retrieval, see the survey [46]. Numerous other nonconvex approaches to phase retrieval exist, which rely on different problem formulations; for example, [13, 19, 66].

Example 2.5 (Robust PCA)**.** In robust principal component analysis, one seeks to identify sparse corruptions of a low-rank matrix [12, 18]. One typical example is image deconvolution, where the low-rank structure models the background of an image while the sparse corruption models the foreground. Formally, given a $m \times n$ matrix M, the goal is to find a decomposition $M = L+S$, where L is low rank and S is sparse. A common relaxation of the problem is

$$\min_{U\in\mathbb{R}^{m\times r}, V\in\mathbb{R}^{n\times r}} \ \|UV^T - M\|_1,$$

where r is the target rank. As is common, the entrywise ℓ_1 norm encourages a sparse residual $UV^T - M$.

Example 2.6 (Censored $\mathbb{Z}_2$ synchronization)**.** A synchronization problem over a graph is to estimate group elements $g_1, \ldots, g_n$ from pairwise products $g_i g_j^{-1}$ over a set of edges $ij \in E$. For a list of applications of such problems see c [1, 5, 65] and references therein. A simple instance is $\mathbb{Z}_2$ synchronization, corresponding to the group on two elements $\{-1, +1\}$. The popular problem of detecting communities in a network, within the binary stochastic block model, can be modeled by using $\mathbb{Z}_2$ synchronization.

Formally, given a partially observed matrix M, the goal is to recover a vector $\theta \in \{\pm 1\}^d$, satisfying $M_{ij} \approx \theta_i\theta_j$ for all $ij \in E$. When the entries of M are corrupted by adversarial sign flips, one can postulate the following formulation:

$$\min_{\theta\in\mathbb{R}^d} \ \|P_E(\theta\theta^T - M)\|_1,$$

where the operator P_E records the entries indexed by the edge set E. Clearly, this is again an instance of (2).

3 Proximally guided subgradient method

As the first example of contemporary applications of the proximal point method, consider the problem of minimizing the expectation:[2]

$$\min_{x\in\mathbb{R}^d} \ F(x) = \mathbb{E}_\zeta f(x, \zeta).$$

Here, ζ is a random variable, and the only access to F is by sampling ζ. It is difficult to overstate the importance of this problem class in large-scale optimization; see, for example, [6, 9].

When the problem is convex, the stochastic subgradient method [49, 58, 60] has strong theoretical guarantees and is often the method of choice. In contrast, when applied to nonsmooth and nonconvex problems, the behavior of the method is poorly understood. The recent paper [24] shows how to use the proximal point method to guide the subgradient iterates in this broader setting, with rigorous guarantees.

Henceforth, assume that the function $x \mapsto f(x, \zeta)$ is ρ-weakly convex and L-Lipschitz for each ζ. Davis and Grimmer [24] proposed the scheme outlined in Algorithm 1.

Algorithm 1: Proximally guided stochastic subgradient method

Data: $x_0 \in \mathbb{R}^d$, $\{j_t\} \subset \mathbb{N}$, $\{\alpha_j\} \subset \mathbb{R}_{++}$
for $t=0,\ldots,T$ **do**
  Set $y_0 = x_t$;
  for $j = 0, \ldots, j_t - 2$ **do**
    Sample ζ and choose
    $v_j \in \partial\Big(f(\cdot, \zeta) + \rho\|\cdot - x_t\|^2\Big)(y_j)$;
    $y_{j+1} = y_j - \alpha_j v_j$
  end
  $x_{t+1} = \frac{1}{j_t}\sum_{j=0}^{j_t-1} y_j$
end

The method proceeds by applying a proximal point method with each subproblem approximately solved by a stochastic subgradient method. The intuition is that each proximal subproblem is $\rho/2$-strongly convex and therefore according to well-known results (e.g., [38, 40, 41, 59]), the stochastic subgradient method should converge at the rate $O(\frac{1}{T})$ on the subproblem, in expectation. This intuition is not quite correct because the objective function of the subproblem is not globally Lipschitz—a key assumption for the $O(\frac{1}{T})$ rate. Nonetheless, the authors show that warm-starting the subgradient method for each proximal subproblem with the current proximal iterate corrects this issue, yielding a favorable guarantee [24, Theorem 1].

To describe the rate of convergence, set $j_t = t + \lceil 648\log(648)\rceil$ and $\alpha_j = \frac{2}{\rho(j+49)}$ in Algorithm 1. Then the scheme will generate an iterate x satisfying

$$\mathbb{E}_\zeta[\|\nabla F_{2\rho}(x)\|^2] \le \varepsilon$$

[2] For simplicity of the exposition, the minimization problem is unconstrained. Simple constraints can be accommodated by using a projection operation.

after at most

$$O\left(\frac{\rho^2(F(x_0)-\min F)^2}{\varepsilon^2}+\frac{L^4\log^4(\varepsilon^{-1})}{\varepsilon^2}\right) \tag{4}$$

subgradient evaluations. This rate agrees with analogous guarantees for stochastic gradient methods for smooth nonconvex functions [36]. Note that convex constraints on x can be easily incorporated into Algorithm 1 by introducing a nearest-point projection in the definition of y_{j+1}.

An interesting and long-standing open question remained: what is the convergence rate of the basic stochastic subgradient method applied directly on the function F (Algorithm 2)?

Algorithm 2: Direct stochastic subgradient method

Data: $x_0 \in \mathbb{R}^d$, $\gamma > 0$
for t=0,...,T **do**
  Sample ζ and choose $v_t \in \partial f(\cdot,\zeta)(x_t)$;
  $x_{t+1} = x_t - \frac{\gamma}{\sqrt{T+1}} v_t$
end

The recent work [22] provided a definitive answer, establishing the following complexity guarantee for Algorithm 2:

$$O\left(\frac{(F(x_0)-\min F)^2}{\gamma^2\varepsilon^2}+\frac{\rho^2\gamma^2L^4}{\varepsilon^2}\right). \tag{5}$$

Here, $\gamma > 0$ is a tuning parameter; for example, setting $\gamma := 1/\rho$ yields the same estimate as (4) up to log factors. The complexity guarantee (5) is somewhat surprising, since neither the Moreau envelope nor the proximal map appear in the definition of Algorithm 2. Nonetheless, the convergence analysis fundamentally relies on using the Moreau envelope as the potential function to monitor along the iterate sequence. In the deterministic setting, one noteworthy advantage of Algorithm 1 is an appealing stopping criterion based on step-size, absent from Algorithm 2.

4 Prox-linear algorithm

For well-structured weakly convex problems, one can hope for faster numerical methods than the subgradient scheme. In this section, I will focus on the composite problem class (2). To simplify the exposition, I will assume $L = 1$, which can always be arranged by rescaling.

Since composite functions are weakly convex, one could apply the proximal point method directly, while setting the parameter $\nu \le \beta^{-1}$. Even though the proximal subproblems are strongly convex, they are not in a form that is most amenable to convex optimization techniques. Indeed, most convex optimization algorithms are designed for minimizing a sum of a convex function and a composition of a convex function with a *linear* map. This observation suggests introducing the following modification to the proximal point algorithm. Given a current iterate x_t, the *prox-linear method* sets

$$x_{t+1} = \operatorname*{argmin}_x \{F(x;x_t) + \tfrac{\beta}{2}\|x-x_t\|^2\},$$

where $F(x;y)$ is the local convex model

$$F(x;y) := g(x) + h\left(c(y)+\nabla c(y)(x-y)\right).$$

In other words, each proximal subproblem is approximated by linearizing the smooth map c at the current iterate x_t.

The main advantage is that each subproblem is now a sum of a strongly convex function and a composition of a Lipschitz convex function with a linear map. A variety of methods utilizing this structure can be formally applied, for example, smoothing [53], saddle-point [17, 48], and interior-point algorithms [51,68]. Which of these methods is practical depends on the specifics of the problem, such as the size and the cost of matrix-vector multiplications.

Note that in the simplest setting of additive composite problems (Example 2.1), the prox-linear method reduces to the popular proximal gradient algorithm or ISTA [8]. For nonlinear least squares, the prox-linear method is a close variant of Gauss-Newton.

Recall that the step size of the proximal point method provides a convenient stopping criterion, since it directly relates to the gradient of the Moreau envelope—a smooth approximation of the objective function. Is there such an interpretation for the prox-linear method? This question is central because termination criteria are used not only to stop the method but also to judge its efficiency and to compare against competing methods.

The answer is yes. Even though one cannot evaluate the gradient $\|\nabla F_{\frac{1}{2\beta}}\|$ directly, the scaled step-size of the prox-linear method

$$\mathcal{G}(x) := \beta(x_{t+1}-x_t)$$

is a good surrogate [31, Theorem 4.5]:

$$\tfrac{1}{4}\|\nabla F_{\frac{1}{2\beta}}(x)\| \le \|\mathcal{G}(x)\| \le 3\|\nabla F_{\frac{1}{2\beta}}(x)\|.$$

In particular, the prox-linear method will find a point x satisfying $\|\nabla F_{\frac{1}{2\beta}}(x)\|^2 \le \varepsilon$ after at most $O\left(\frac{\beta(F(x_0)-\inf F)}{\varepsilon}\right)$ iterations. In the simplest setting when $g = 0$ and $h(t) = t$, this rate reduces to the well-known convergence guarantee of gradient descent, which is black-box optimal for C^1-smooth nonconvex optimization [15].

A number of improvements to the basic prox-linear method were recently proposed. Cartis et al. [16] discuss trust-region variants and their complexity guarantees. Duchi and Ruan [33] propose a stochastic extension of the scheme and prove almost sure convergence, while the convergence rate for the stochastic prox-linear method is proved in [21]. In [31], the authors discuss overall complexity guarantees of the prox-linear method when the convex subproblems can be solved only by first-order methods, and propose an inertial variant of the scheme whose convergence guarantees automatically adapt to the near-convexity of the problem.

Local rapid convergence

Under typical regularity conditions, the prox-linear method exhibits the same types of rapid convergence guarantees as the proximal point method. I will illustrate with two intuitive and widely used regularity conditions, yielding local linear and quadratic convergence, respectively.

Definition 2 ([57]). A local minimizer $\bar{x}$ of F is *α-tilt stable* if there exists $r > 0$ such that the solution map

$$M : v \mapsto \operatorname*{argmin}_{x \in B_r(\bar{x})} \{F(x) - \langle v, x \rangle\}$$

is $1/\alpha$-Lipschitz around 0 with $M(0) = \bar{x}$.

This condition might seem unfamiliar to convex optimization specialists. Although not obvious, tilt stability is equivalent to a uniform quadratic growth property and a subtle localization of strong convexity of F. See [28] or [30] for more details on these equivalences. Under the tilt stability assumption, the prox-linear method initialized sufficiently close to $\bar{x}$ produces iterates that converge at a linear rate of $1 - \alpha/\beta$.

The second regularity condition models sharp growth of the function around the minimizer. Let S be the set of all stationary points of F, meaning x lies in S if and only if the directional derivative $F'(x; v)$ is nonnegative in every direction $v \in \mathbb{R}^d$.

Definition 3 ([10]). A local minimizer $\bar{x}$ of F is *sharp* if there exist $\alpha > 0$ and a neighborhood $\mathcal{X}$ of $\bar{x}$ such that

$$F(x) \geq F(\mathrm{proj}_S(x)) + c \cdot \mathrm{dist}(x, S) \qquad \forall x \in \mathcal{X}.$$

Under the sharpness condition, the prox-linear method initialized sufficiently close to $\bar{x}$ produces iterates that converge quadratically.

For well-structured problems, one can hope to justify the two regularity conditions under statistical assumptions. The work of Duchi and Ruan on the phase retrieval problem [32] is an interesting recent example. Under mild statistical assumptions on the data-generating mechanism, sharpness is ensured with high probability. Therefore the prox-linear method (and even a subgradient method [23]) converges rapidly, when initialized within a constant relative distance of an optimal solution.

5 Catalyst acceleration

The final example concerns inertial acceleration in convex optimization. Setting the groundwork, consider a μ-strongly convex function f with a β-Lipschitz gradient map $x \mapsto \nabla f(x)$. Classically, gradient descent will find a point x satisfying $f(x) - \min f < \varepsilon$ after at most

$$O\left(\frac{\beta}{\mu} \ln(1/\varepsilon)\right)$$

iterations. Accelerated gradient methods, beginning with Nesterov [52], equip the gradient descent method with an inertial correction. Such methods have the much lower complexity guarantee

$$O\left(\sqrt{\frac{\beta}{\mu}} \ln(1/\varepsilon)\right),$$

which is optimal within the first-order oracle model of computation [50].

One naturally may ask which other methods, aside from gradient descent, can be "accelerated." For example, one may wish to accelerate coordinate descent or so-called variance-reduced methods for finite-sum problems; I will comment on the latter problem class shortly.

One appealing strategy relies on the proximal point method. Güler in [37] showed that the proximal point method itself can be equipped with inertial steps leading to improved convergence guarantees. Building on this work, Lin, Mairal, and Harchaoui [45] explained how to derive the *total* complexity guarantees for an inexact accelerated proximal point method that take into account the cost of applying an arbitrary linearly convergent algorithm $\mathcal{M}$ to the subproblems. Their *Catalyst acceleration* framework is summarized in Algorithm 3.

Algorithm 3: Catalyst acceleration

Data: $x_0 \in \mathbb{R}^d$, $\kappa > 0$, algorithm $\mathcal{M}$
Set $q = \mu/(\mu + \kappa)$, $\alpha_0 = \sqrt{q}$, and $y_0 = x_0$;
for *t=1,…,T* **do**
Use $\mathcal{M}$ to approximately solve:

$$x_t \approx \operatorname*{argmin}_{x \in \mathbb{R}^d} \left\{F(x) + \frac{\kappa}{2}\|x - y_{t-1}\|^2\right\}. \qquad (6)$$

Compute $\alpha_t \in (0, 1)$ from the equation

$$\alpha_t^2 = (1 - \alpha_t)\alpha_{t-1}^2 + q\alpha_t.$$

Compute:

$$\beta_t = \frac{\alpha_{t-1}(1 - \alpha_{t-1})}{\alpha_{t-1}^2 + \alpha_t},$$
$$y_t = x_t + \beta_t(x_t - x_{t-1}).$$

end

To state the guarantees of this method, suppose that $\mathcal{M}$ converges on the proximal subproblem in function value at a linear rate $1 - \tau \in (0, 1)$. Then a simple termination policy on the subproblems (6) yields an algorithm with overall complexity

$$\widetilde{O}\left(\frac{\sqrt{\mu + \kappa}}{\tau\sqrt{\mu}} \ln(1/\varepsilon)\right). \qquad (7)$$

That is, the expression (7) describes the maximal number of iterations of $\mathcal{M}$ used by Algorithm 3 until it finds a point x satisfying $f(x) - \inf f \leq \varepsilon$. Typically τ depends on κ; therefore the best choice of κ is the one that minimizes the ratio $\frac{\sqrt{\mu+\kappa}}{\tau\sqrt{\mu}}$.

The main motivation for the Catalyst framework, and its most potent application, is the regularized empirical risk minimization (ERM) problem:

$$\min_{x\in\mathbb{R}^d} f(x) := \frac{1}{m}\sum_{i=1}^{m} f_i(x) + g(x).$$

Such large-finite sum problems are ubiquitous in machine learning and high-dimensional statistics, where each function f_i typically models a misfit between predicted and observed data, while g promotes some low-dimensional structure on x, such as sparsity or low rank.

Assume that f is μ-strongly convex and each individual f_i is C^1-smooth with β-Lipschitz gradient. Since m is assumed to be huge, the complexity of numerical methods is best measured in terms of the total number of individual gradient evaluations ∇f_i. In particular, fast gradient methods have the worst-case complexity,

$$O\left(m\sqrt{\frac{\beta}{\mu}}\ln(1/\varepsilon)\right),$$

since each iteration requires evaluation of all the individual gradients $\{\nabla f_i(x)\}_{i=1}^m$. Variance-reduced algorithms, such as SAG [62], SAGA [26], SDCA [63], SMART [20], SVRG [39, 69], FINITO [27], and MISO [45, 47], aim to improve the dependence on m. In their raw form, all these methods exhibit a similar complexity,

$$O\left(\left(m+\frac{\beta}{\mu}\right)\ln(1/\varepsilon)\right),$$

in expectation, and differ only in storage requirements and in whether one needs to know explicitly the strong convexity constant.

A long-standing open question was to determine whether the dependence on β/μ can be improved. This is not possible in full generality, and instead one should expect a rate of the form

$$O\left(\left(m+\sqrt{m\frac{\beta}{\mu}}\right)\ln(1/\varepsilon)\right). \tag{8}$$

Indeed, such a rate would be optimal in certain regimes [2, 4, 42, 67]. Note that the complexity (8) is beneficial only in the setting $m < \beta/\mu$.

Early examples for specific algorithms are the accelerated SDCA [64], APPA [35], and RPDG [42].[3] The accelerated SDCA and APPA, in particular, use a specialized proximal point construction.[4] Catalyst generic acceleration allows all of the variance-reduced methods above to be accelerated in a single conceptually transparent framework. Note that the first direct accelerated variance-reduced methods for ERM problems were recently proposed in [3, 25].

In contrast to the convex setting, the role of inertia for nonconvex problems is not nearly as well understood. In particular, gradient descent is black-box optimal for C^1-smooth nonconvex minimization [15], and therefore inertia cannot help in the worst case. On the other hand, the recent paper [14] presents a first-order method for minimizing C^2 and C^3 smooth functions that is provably faster than gradient descent. At its core, the algorithm also combines inertia with the proximal point method. For a partial extension of the Catalyst framework to weakly convex problems, see [56].

[3] Here, I am ignoring logarithmic terms in the convergence rate.

[4] The accelerated SDCA was the motivation for the Catalyst framework, while APPA appeared concurrently with Catalyst.

6 Conclusion

The proximal point method has long been ingrained in the foundations of optimization. Recent progress in large-scale computing has shown that the proximal point method not only is conceptual but can guide methodology. Although direct methods are usually preferable, proximally guided algorithms can be equally effective and often lead to more easily interpretable numerical methods. In this article, I outlined three examples of this viewpoint, where the proximal point method guides both the design and the analysis of numerical methods.

Acknowledgments. I thank Damek Davis, John Duchi, and Zaid Harchaoui for their helpful comments on an early draft of the article. This research was supported by the AFOSR YIP award FA9550-15-1-0237 and by the NSF DMS 1651851 and CCF 1740551 awards.

REFERENCES

[1] E. Abbe, A.S. Bandeira, A. Bracher, and A. Singer. Decoding binary node labels from censored edge measurements: phase transition and efficient recovery. *IEEE Trans. Network Sci. Eng.*, 1(1):10–22, 2014.

[2] A. Agarwal and L. Bottou. A lower bound for the optimization of finite sums. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 78–86, 2015.

[3] Z. Allen-Zhu. Katyusha: The first direct acceleration of stochastic gradient methods. In *Proc. STOC 2017*, pages 1200–1205, 2017.

[4] Y. Arjevani. Limitations on variance-reduction and acceleration schemes for finite sums optimization. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3543–3552. Curran Associates, Inc., 2017.

[5] A.S. Bandeira, N. Boumal, and V. Voroninski. On the low-rank approach for semidefinite programs arising in synchronization and community detection. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, June 23-26, 2016*, pages 361–382, 2016.

[6] P.L. Bartlett, M.I. Jordan, and J.D. McAuliffe. Convexity, classification, and risk bounds. *J. Amer. Statist. Assoc.*, 101(473):138–156, 2006.

[7] H.H. Bauschke and P.L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, second edition, 2017. With a foreword by Hédy Attouch.

[8] A. Beck and M. Teboulle. Smoothing and first order methods: a unified framework. *SIAM J. Optim.*, 22(2):557–580, 2012.

[9] L. Bottou and O. Bousquet. The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems*, pages 161–168, 2008.

[10] J.V. Burke and M.C. Ferris. Weak sharp minima in mathematical programming. *SIAM J. Control Optim.*, 31(5):1340–1359, 1993.

[11] J.V. Burke and M.C. Ferris. A Gauss-Newton method for convex composite optimization. *Math. Program.*, 71(2, Ser. A):179–194, 1995.

[12] E.J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):Art. 11, 37, 2011.

[13] E.J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: theory and algorithms. *IEEE Trans. Inform. Theory*, 61(4):1985–2007, 2015.

[14] Y. Carmon, J.C. Duchi, O. Hinder, and A. Sidford. "Convex until proven guilty": Dimension-free acceleration of gradient descent on non-convex functions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 654–663, 2017.

[15] Y. Carmon, J.C. Duchi, O. Hinder, and A. Sidford. Lower bounds for finding stationary points I. *Preprint arXiv:1710.11606*, 2017.

[16] C. Cartis, N.I.M. Gould, and P.L. Toint. On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. *SIAM J. Optim.*, 21(4):1721–1739, 2011.

[17] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[18] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.*, 21(2):572–596, 2011.

[19] Y. Chen and E.J. Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Comm. Pure Appl. Math.*, 70(5):822–883, 2017.

[20] D. Davis. SMART: The stochastic monotone aggregated root-finding algorithm. *Preprint arXiv:1601.00698*, 2016.

[21] D. Davis and D. Drusvyatskiy. Stochastic model-based minimization of weakly convex functions. *Preprint arXiv:1803.06523*, 2018.

[22] D. Davis and D. Drusvyatskiy. Stochastic subgradient method converges at the rate $O(k^{-1/4})$ on weakly convex functions. *Preprint arXiv:1802.02988*, 2018.

[23] D. Davis, D. Drusvyatskiy, and C. Paquette. The nonsmooth landscape of phase retrieval. *Preprint arXiv:1711.03247*, 2017.

[24] D. Davis and B. Grimmer. Proximally guided stochastic subgradient method for nonsmooth, nonconvex problems. *Preprint arXiv:1707.03505*, 2017.

[25] A. Defazio. A simple practical accelerated method for finite sums. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 676–684. Curran Associates, Inc., 2016.

[26] A. Defazio, F. Bach, and S. Lacoste-Julien. SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1646–1654. Curran Associates, Inc., 2014.

[27] A. Defazio, J. Domke, and T.S. Caetano. Finito: A faster, permutable incremental gradient method for big data problems. In *ICML*, pages 1125–1133, 2014.

[28] D. Drusvyatskiy and A.S. Lewis. Tilt stability, uniform quadratic growth, and strong metric regularity of the subdifferential. *SIAM J. Optim.*, 23(1):256–267, 2013.

[29] D. Drusvyatskiy and A.S. Lewis. Error bounds, quadratic growth, and linear convergence of proximal methods. *Math. Oper. Res.*, 2018. To appear, doi:10.1287/moor.2017.0889.

[30] D. Drusvyatskiy, B.S. Mordukhovich, and T.T.A. Nghia. Second-order growth, tilt-stability, and metric regularity of the subdifferential. *J. Convex Anal.*, 21(4):1165–1192, 2014.

[31] D. Drusvyatskiy and C. Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Preprint arXiv:1605.00125*, 2016.

[32] J.C. Duchi and F. Ruan. Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval. *Preprint arXiv:1705.02356*, 2017.

[33] J.C. Duchi and F. Ruan. Stochastic methods for composite optimization problems. *Preprint arXiv:1703.08570*, 2017.

[34] Y.C. Eldar and S. Mendelson. Phase retrieval: stability and recovery guarantees. *Appl. Comput. Harmon. Anal.*, 36(3):473–494, 2014.

[35] R. Frostig, R. Ge, S.M. Kakade, and A. Sidford. Un-regularizing: approximate proximal point and faster stochastic algorithms for empirical risk minimization. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, 2015.

[36] S. Ghadimi and G. Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM J. Optim.*, 23(4):2341–2368, 2013.

[37] O. Güler. New proximal point algorithms for convex minimization. *SIAM J. Optim.*, 2(4):649–664, 1992.

[38] E. Hazan and S. Kale. Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization. In Sham M. Kakade and Ulrike von Luxburg, editors, *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *Proceedings of Machine Learning Research*, pages 421–436, Budapest, Hungary, 09–11 Jun 2011. PMLR.

[39] R. Johnson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, NIPS'13, pages 315–323. Curran Associates Inc., 2013.

[40] A. Juditsky and Y. Nesterov. Deterministic and stochastic primal-dual subgradient algorithms for uniformly convex minimization. *Stoch. Syst.*, 4(1):44–80, 2014.

[41] S. Lacoste-Julien, M. Schmidt, and F. Bach. A simpler approach to obtaining an $O(1/t)$ convergence rate for the projected stochastic subgradient method. *Preprint arXiv:1212.2002*, 2012.

[42] G. Lan and Y. Zhou. An optimal randomized incremental gradient method. *Math. Program.*, Jun 2017.

[43] C. Lemarechal, J.-J. Strodiot, and A. Bihain. On a bundle algorithm for nonsmooth optimization. In *Nonlinear Programming, 4 (Madison, Wis., 1980)*, pages 245–282. Academic Press, New York-London, 1981.

[44] A.S. Lewis and S.J. Wright. A proximal method for composite minimization. *Math. Program.*, 158:501–546, 2016.

[45] H. Lin, J. Mairal, and Z. Harchaoui. A universal catalyst for first-order optimization. In *Advances in Neural Information Processing Systems*, pages 3366–3374, 2015.

[46] R. Luke. Phase retrieval, what's new? *SIAG/OPT Views and News*, 25(1), 2017.

[47] J. Mairal. Incremental majorization-minimization optimization with application to large-scale machine learning. *SIAM J. Optim.*, 25(2):829–855, 2015.

[48] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.*, 15(1):229–251, 2004.

[49] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.*, 19(4):1574–1609, 2008.

[50] A.S. Nemirovsky and D.B. Yudin. *Problem Complexity and Method Efficiency in Optimization.* A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1983. Translated from the Russian and with a preface by E. R. Dawson, Wiley-Interscience Series in Discrete Mathematics.

[51] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13 of *SIAM Studies in Applied Mathematics*. SIAM, 1994.

[52] Yu. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR*, 269(3):543–547, 1983.

[53] Yu. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1, Ser. A):127–152, 2005.

[54] Yu. Nesterov. Modified Gauss-Newton scheme with worst case guarantees for global performance. *Optim. Methods Softw.*, 22(3):469–483, 2007.

[55] Yu. Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, 140(1, Ser. B):125–161, 2013.

[56] C. Paquette, H. Lin, D. Drusvyatskiy, J. Mairal, and Z. Harchaoui. Catalyst acceleration for gradient-based non-convex optimization. *Preprint arXiv:1703.10993*, 2017.

[57] R.A. Poliquin and R.T. Rockafellar. Tilt stability of a local minimum. *SIAM J. Optim.*, 8(2):287–299, 1998.

[58] B.T. Polyak and A.B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4):838–855, 1992.

[59] A. Rakhlin, O. Shamir, and K. Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pages 1571–1578. Omnipress, 2012.

[60] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.

[61] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.

[62] M. Schmidt, N. Le Roux, and F. Bach. Minimizing finite sums with the stochastic average gradient. *Math. Program.*, 162(1-2):83–112, March 2017.

[63] S. Shalev-Shwartz and T. Zhang. Proximal stochastic dual coordinate ascent. *Preprint arXiv:1211.2717*, 2012.

[64] S. Shalev-Shwartz and T. Zhang. Accelerated proximal stochastic dual coordinate ascent for regularized loss minimization. *Math. Program.*, 156(1-2):105–145, 2016.

[65] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, 30(1):20–36, 2011.

[66] J. Sun, Q. Qu, and J. Wright. A geometric analysis of phase retrieval. *Found. Comp. Math.*, 2018. To appear, doi 10.1007/s10208-017-9365-9.

[67] B.E. Woodworth and N. Srebro. Tight complexity bounds for optimizing composite objectives. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 3639–3647. Curran Associates, Inc., 2016.

[68] S.J. Wright. *Primal-Dual Interior-Point Methods.* SIAM, 1997.

[69] L. Xiao and T. Zhang. A proximal stochastic gradient method with progressive variance reduction. *SIAM J. Optim.*, 24(4):2057–2075, 2014.

Bulletin

1 Event Announcements

1.1 2018 DIMACS/TRIPODS/MOPTA Conference

This year the annual MOPTA conference will be combined with an NSF-TRIPODS sponsored three day summer school for doctoral students and the NSF-DIMACS sponsored workshop on Optimization in Machine Learning which is a part of the DIMACS/Simons Collaboration on Bridging Continuous and Discrete Optimization.

The summer school will be held during the three days preceding the conference and is designed for doctoral students interested in improving their theoretical and practical skills related to optimization approaches in machine learning. The DIMACS sponsored workshop will bring together invited lectures by top experts in the field as well as contributed poster presentations.

The MOPTA part of the conference this year will include a variety on exciting new developments from different optimization areas with a special focus on applications in energy. It will aim to bring together researchers from both theoretical and applied communities who do not usually have the chance to interact in the framework of a medium-scale event.

Plenary speakers for MOPTA:
Daniel Bienstock (Columbia), Marija Ilic (MIT), Andrea Lodi (UMontreal), David Morton (Northwestern).
Plenary speakers for DIMACS/TRIPODS:
Peter Bartlett (Berkeley), John Duchi (Stanford), Suvrit Sra (MIT), Kilian Weinberger (Cornell), Stephen Wright (Wisconsin).

More details are available on the conference website: http://coral.ie.lehigh.edu/~mopta/.

1.2 MINLP summer school at Trier University

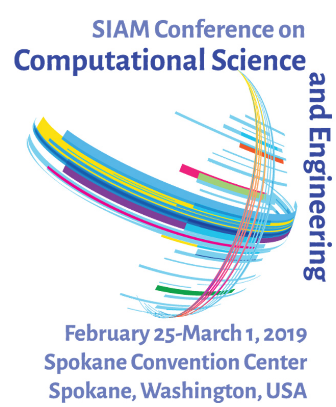
The research training group Algorithmic Optimization (ALOP) at Trier University, Germany, organizes a summer school on "Mixed-Integer Nonlinear Programming" from August 13-16, 2018 (https://alop.uni-trier.de/MINLP/).

The lectures will be given by: Oliver Bastert & Zsolt Csizmadia (FICO, Xpress Optimization), Christoph Buchheim (Technical University of Dortmund), Sven Leyffer (Argonne National Laboratory), and Jeff Linderoth (University of Wisconsin-Madison).

MINLP combines the complexity of Integer Programming with the challenges of Nonlinear Optimization. The summer school aims at giving an overview of the many different aspects of this exciting field of mathematical optimization for MSc and PhD students in mathematics, computer science and related fields.

ALOP is accepting applications for travel funding for students attending the summer school. The travel support application deadline is May 30, 2018.

1.3 2019 SIAM Conference on Computational Science and Engineering



The SIAM Conference on Computational Science and Engineering (CSE19) seeks to enable in-depth technical discussions on a wide variety of major computational efforts on large-scale problems in science and engineering, foster the interdisciplinary culture required to meet these large-scale challenges, and promote the training of the next generation of computational scientists.

Themes for this conference are as follows:

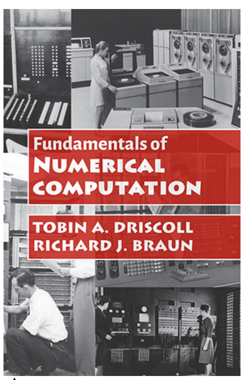
- Computational science and machine learning
- Statistical modeling, methods, and computation
- Multiscale, multiphysics, and multilevel methods
- High performance software: packages and design
- Algorithms at extreme scales
- Tensor Computations
- High-order methods, novel discretizations, and scalable solvers
- Data science, analytics, and visualization
- Applications in science, engineering, and industry
- Biological and biomedical computations
- Scientific simulation and uncertainty
- Numerical optimization: methods and applications
- Reduced order modeling
- Emerging trends in CS&E education and training

Deadlines (Midnight Eastern Time)
July 25: Minisymposium Proposal Submissions
August 22: Contributed Lecture, Poster, and Minisymposium Presentation Abstracts

More details are available on the conference website: http://www.siam.org/meetings/cse19/.

2 Book Announcements

2.1 Fundamentals of Numerical Computation



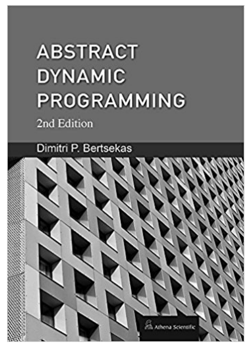
By Tobin A. Driscoll and Richard J. Braun
Publisher: SIAM
ISBN: 978-1-611975-07-9, xxx + 553 pages
Published: 2017
http://bookstore.siam.org/ot154/

About the book: *Fundamentals of Numerical Computation* is an advanced undergraduate-level introduction to the mathematics and use of algorithms for the fundamental problems of numerical computation: linear algebra, finding roots, approximating data and functions, and solving differential equations. The book is organized with simpler methods in the first half and more advanced methods in the second half, allowing use for either a single course or a sequence of two courses. The authors take readers from basic to advanced methods, illustrating them with over 200 self-contained MATLAB functions and examples designed for those with no prior MATLAB experience. Although the text provides many examples, exercises, and illustrations, the aim of the authors is not to provide a cookbook per se, but rather an exploration of the principles of cooking.

Professors Driscoll and Braun have developed an online resource that includes well-tested materials related to every chapter. Among these materials are lecture-related slides and videos, ideas for student projects, laboratory exercises, computational examples and scripts, and all the functions presented in the book.

Audience: Fundamentals of Numerical Computation is intended for advanced undergraduates in math, applied math, engineering, or science disciplines, as well as for researchers and professionals looking for an introduction to a subject they missed or overlooked in their education.

2.2 Abstract Dynamic Programming, 2nd Ed.



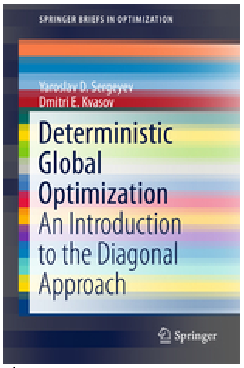
By Dimitri P. Bertsekas
Publisher: Athena Scientific
ISBN: 978-1-886529-46-5, 360 pages
Published: February 2018
http://www.athenasc.com/abstractdp.html

About the book: The book provides a synthesis of old research on the foundations of dynamic programming, the modern theory of approximate dynamic programming/reinforcement learning, and new research on semicontractive models, a broad generalization of shortest path-type problems.

It aims at a unified and economical development of the core theory and algorithms of total cost sequential decision problems, based on the strong connections of the subject with fixed point theory. The analysis focuses on the abstract mapping that underlies dynamic programming and defines the mathematical character of the associated problem. The discussion centers on two fundamental properties that this mapping may have: monotonicity and (weighted sup-norm) contraction. It turns out that the nature of the analytical and algorithmic DP theory is determined primarily by the presence or absence of these two properties, and the rest of the problem's structure is largely inconsequential. New research is focused on two areas: 1) The ramifications of these properties in the context of algorithms for approximate dynamic programming, and 2) The new class of semicontractive models, exemplified by shortest path problems, where some but not all policies are contractive/terminating.

The new edition aims primarily to amplify the presentation of the semicontractive models of Chapter 3 and Chapter 4 of the first (2013) edition, and to supplement it with a broad spectrum of research results that I obtained and published in journals and reports since the first edition was written. As a result, the size of this material more than doubled, and the size of the book increased by nearly 40.

2.3 Deterministic Global Optimization: An Introduction to the Diagonal Approach



By Yaroslav D. Sergeyev and Dmitri E. Kvasov
Publisher: Springer-Verlag New York
Series: Springer Briefs in Optimization
ISBN: 978-1-4939-7197-8, x + 136 pages
Published: 2017
http://www.springer.com/us/book/9781493971978

About the book: This book begins with a concentrated introduction into deterministic global optimization and moves forward to present new original results from the authors who are well known experts in the field. Multiextremal continuous problems that have an unknown structure with Lipschitz objective functions and functions having the first Lipschitz derivatives defined over hyperintervals are studied. A class of algorithms using several Lipschitz constants is introduced which has its origins in the DIRECT (DIviding RECTangles) method. This new class is based on an efficient strategy that is applied for the search domain partitioning. In addition a survey on derivative free methods and methods using the first derivatives is given for both one- dimensional and multi-dimensional cases. Non-smooth and smooth minorants and acceleration techniques that can speed up several classes of global optimization methods with examples of applications and problems arising in numerical testing of global optimization algorithms are discussed. Theoretical considerations are illustrated through engineering applications. Extensive numerical testing of algorithms described in this book stretches the likelihood of establishing a link between mathematicians and practitioners. The authors conclude by describing applications and a generator of random classes of test functions with known local and global minima that is used in more than 40 countries of the world.

Audience: This book serves as a starting point for students, researchers, engineers, and other professionals in operations research, management science, computer science, engineering, economics, environmental sciences, industrial and applied mathematics to obtain an overview of deterministic global optimization.

3 Other Announcements

3.1 2018 SIAM Fellows Announced

Each year, SIAM designates as Fellows of the society those who have made outstanding contributions to the fields of applied mathematics and computational science. This year, 28 members of the community were selected for this distinction.

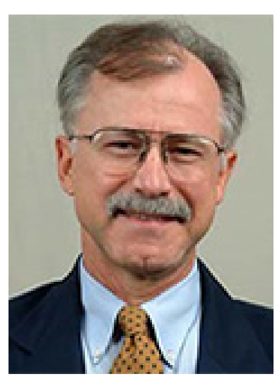
These new Fellows include five members of the SIAG, whose citations are included below. Full details on the SIAM Fellow program can be found at http://www.siam.org/prizes/fellows/index.php. Congratulations to all the new Fellows!

Helen Moore
AstraZeneca
For impactful industrial application of mathematical modeling in oncology, immunology, and virology. For mentoring, teaching, and leadership.

Pablo A. Parrilo
Massachusetts Institute of Technology
For foundational contributions to algebraic methods in optimization and engineering.

Tamás Terlaky
Lehigh University *For fundamental and sustained contributions to the theory and practice of optimization, and for exemplary service to the optimization community.*

Kim-Chuan Toh
National University of Singapore
For his contributions to the development of algorithms and software for semidefinite programming and, more generally, conic programming.



Homer F. Walker
Worcester Polytechnic Institute
For contributions to theory and software of iterative methods for nonlinear systems and optimization, as well as application of these methods to scientific simulations.

3.2 MPC "Best Paper of the Year" 2017

Congratulations to Diego Pecin, Artur Pessoa, Marcus Poggi, and Eduardo Uchoa for winning "Best Paper of the Year" for 2017 from Mathematical Programming Computation (MPC). Their paper entitled "Improved branch-cut-and-price for capacitated vehicle routing" (*Mathematical Programming Computation*, Volume 9, Issue 1, pp. 61–100, March 2017) was chosen from among all papers that were in print in MPC in 2017. This is the inaugural prize for MPC's best paper of the year.

3.3 James H. Wilkinson Prize for Numerical Software - Call for Entries

Entries for the James H. Wilkinson Prize for Numerical Software are currently begin accepted. The prize is awarded every four years to the authors of an outstanding piece of numerical software. The prize is awarded for an entry that best addresses all phases of the preparation of high-quality numerical software. It is intended to recognize innovative software in scientific computing and to encourage researchers in the earlier stages of their career.

SIAM will award the Wilkinson Prize for Numerical Software at the SIAM Conference on Computational Science and Engineering (CSE19). The award will consist of $3,000 and a plaque. As part of the award, the recipient(s) will be expected to present a lecture at the conference.

Eligibility Criteria: Selection will be based on: clarity of the software implementation and documentation, importance of the application(s) addressed by the software; portability, reliability, efficiency, and usability of the software implementation; clarity and depth of analysis of the algorithms and the software in the accompanying paper; and quality of the test software.

Candidates must have worked in mathematics or science for at most 12 years (full time equivalent) after receiving their PhD as of January 1 of the award year, allowing for breaks in continuity. The prize committee can make exceptions, if in their opinion the candidate is at an equivalent stage in their career. For the 2019 award, a candidate must have received their PhD no earlier than January 1, 2007. The entry deadline is June 1, 2018. Submission instructions can be found at http://www.siam.org/prizes/nominations/nom_wilkinson_ns.php.

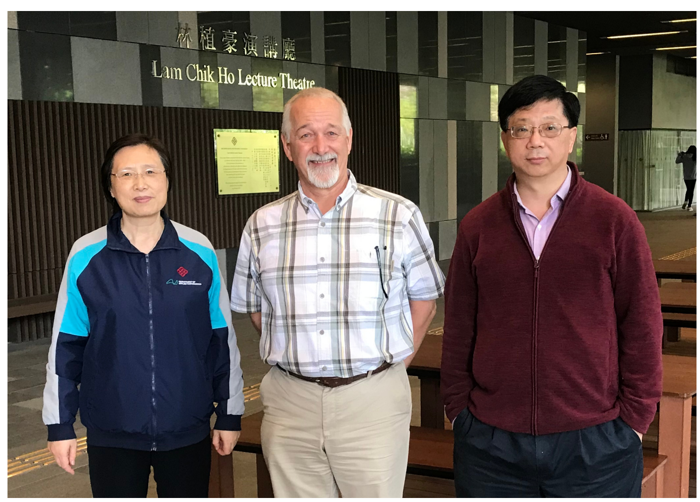
Selection Committee: Jorge Moré (Chair), Argonne National Laboratory; Sven Hammarling, Numerical Algorithms Group Ltd and University of Manchester; Michael Heroux, Sandia National Laboratories; Randall J. LeVeque, University of Washington; Katherine Yelick, Lawrence Berkeley National Laboratory

Chair's Column

The SIAG/OPT leadership worked on a few important projects. SIAG/OPT was asked to organize a series of featured minisymposia for the SIAM Annual Meeting held in Portland, OR, July 9–13. Naturally, Michael Friedlander, our program director, took the lead of this activity. The SIAM Annual meeting includes featured presentations highly relevant to optimization: a plenary talk on Algebraic Vision by Rekha R. Thomas (U. Washington), the John von Neumann Lecture by Charles F. Van Loan, (Cornell U.), and the AWM-SIAM Sonia Kovalevsky Lecture by Eva Tardos (Cornell U.).

The Annual Meeting will also include the induction ceremony of the 2018 class of SIAM Fellows, which includes SIAG/OPT members Helen Moore, Pablo A. Parrilo, Tamás Terlaky, Kim-Chuan Toh, and Homer F. Walker. My sincere congratulations to all Fellows! It is my great honor to join such a distinguished club of colleagues.

Regarding the SIAM Conference on Optimization (OP20), the contract between SIAM and the local organizers is signed, so all is set for a great conference. In March, I visited Hong Kong Polytechnic and the conference facilities, and I also met with the Local Organizing Committee Chair Xiao-jun Chen and Organizing Committee Co-Chair Defeng Sun. I am pleased to report that Xiao-jun secured superb conference facilities. With the exception of the opening ceremony, all sessions will be in a modern building spread only between two levels with easy escalator and stair access. The longest walking distance between two session rooms is less than 3 minutes. There are large lecture theaters for plenary and semi-plenary lectures, as well as an ample supply of smaller breakout rooms. The opening plenary will be in a beautiful and large theater, about 5 minutes walking distance from the central conference place. Hong Kong is a vibrant, wonderful, and multicultural city that offers a wide variety of hotels nearby at various price levels. With Defeng we finalized the OP20's Organizing Committee. Stay tuned for more about the Organizing Committee's activities in the next newsletter.



Xiao-jun Chen, Tamás Terlaky, and Defeng Sun surveying the facilities for the OP20 meeting in Hong Kong.

With the SIAG/OPT leadership team, we have developed a proposal to establish an Early Career prize. We have repeatedly discussed details of the proposal with SIAM's major award committee to ensure that the SIAM Optimization

Early Career prize is consistent with SIAM's rules and analogous prizes of other activity groups, and in the same time serves our activity group best. We certainly will present the inaugural Early Career Prize in Hong Kong at OP20; stay tuned for a call for nominations.

Finally, I wish all of you a wonderful and productive summer! Looking forward to see many of you in Portland at the Annual Meeting.

Tamás Terlaky, SIAG/OPT Chair
Department of Industrial and Systems Engineering, P.C. Rossin College of Engineering and Applied Science, Lehigh University, Bethlehem, PA 18015-1582, USA,
terlaky@lehigh.edu, http://www.lehigh.edu/~tat208

Comments from the Editors

Electronic preprint services (in particular those spanning institutions) continue to see remarkable rises in submissions. As of April, the editors' preference for optimization-centric preprints, Optimization Online (OO) has seen over 6,300 submissions. Submissions to the more general-purpose arXiv exceeded 120,000 per year in 2017. ArXiv submissions in the subject areas of mathematics and computer science now make up half of all submissions; see Fig. A.

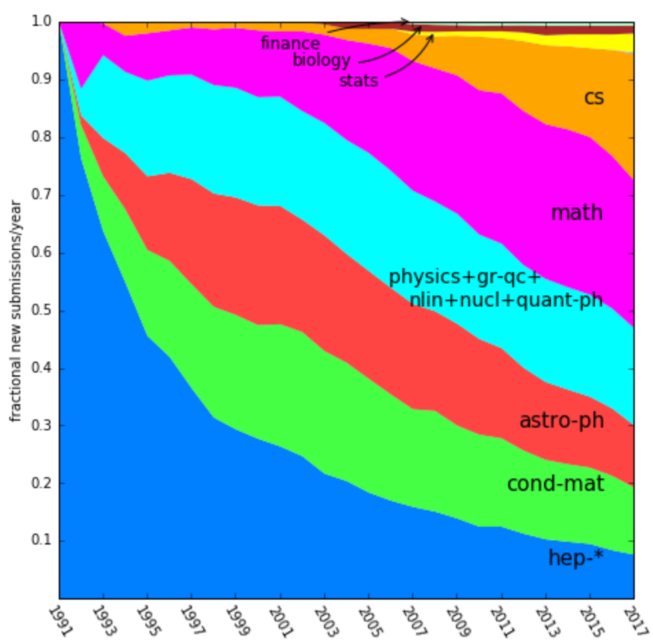


Fig. A: Fraction of arXiv submissions by subject.

We see this firsthand in *Views and News*. In the ten-year period between 2007 and 2016, 1.2% (7 out of 570) of the citations in *Views and News* pointed to an arXiv preprint. The three issues since then clock in at a rate of 10.5% (26 out of 248). This could be editorial bias, a growing gap between electronic preprint and publication appearance (after all, many of the recent nonpreprint cited works also originally appeared on OO or the arXiv), and other factors. Even within mathematics, the growth of optimization and control submissions is steady; see Fig. B.

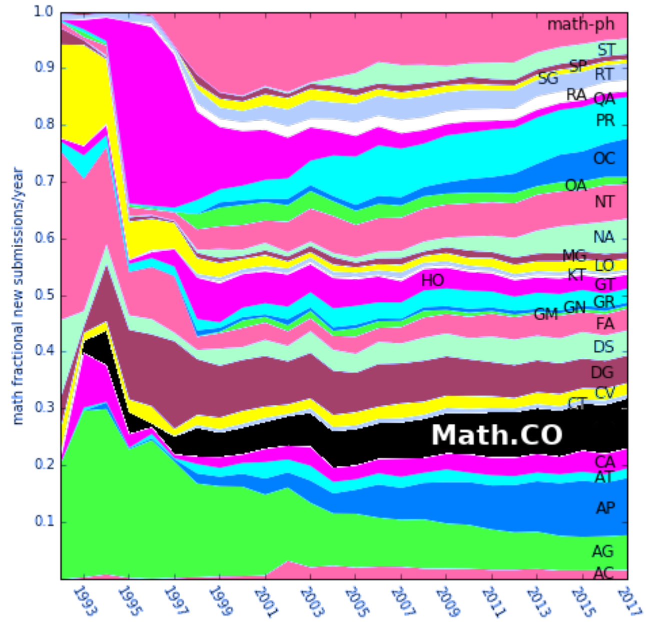


Fig. B: Fraction of mathematics submissions to arXiv by subarea; Optimization and Control subarea highlighted.

We note that the article in this issue is available at https://arxiv.org/abs/1712.06038. What are your thoughts on the growth, effects, and drivers of electronic preprints?

Digital/Printed Newsletters. Many of you have written to opt for an electronic copy of *Views and News*; if you are receiving physical copies and prefer otherwise, please do not hesitate to contact us.

Upcoming Meetings. We are excited to see many of you in a couple of months in Bordeaux at the 2018 International Symposium on Mathematical Programming (ISMP 2018). Before we close out another decade (time really does fly when you're having fun!), the recently announced SIAM CSE19 meeting in Spokane, Washington includes a number of themes of direct interest to our activity group.

As always, we welcome your feedback, emailed directly to us or to siagoptnews@lists.mcs.anl.gov. Suggestions for new issues, comments, and papers are always welcome!

Congratulations to the activity group's chair and all of the new SIAM fellows!

Stefan M. Wild, Editor
Mathematics & Computer Science Div., Argonne National Lab, USA, wild@anl.gov, http://www.mcs.anl.gov/~wild
Jennifer Erway, Editor
Department of Mathematics, Wake Forest University, USA, erwayjb@wfu.edu, http://www.wfu.edu/~erwayjb