J. Math. Anal. Appl. 457 (2018) 1283–1306

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Perspective functions: Proximal calculus and applications in high-dimensional statistics

Patrick L. Combettes [a,*], Christian L. Müller [b]

[a] North Carolina State University, Department of Mathematics, Raleigh, NC 27695-8205, USA
[b] Flatiron Institute, Simons Foundation, New York, NY 10010, USA

ARTICLE INFO

Article history:
Received 12 October 2016
Available online 15 December 2016
Submitted by H. Frankowska

Keywords:
Convex function
Perspective function
Proximal algorithm
Proximity operator
Statistics

ABSTRACT

Perspective functions arise explicitly or implicitly in various forms in applied mathematics and in statistical data analysis. To date, no systematic strategy is available to solve the associated, typically nonsmooth, optimization problems. In this paper, we fill this gap by showing that proximal methods provide an efficient framework to model and solve problems involving perspective functions. We study the construction of the proximity operator of a perspective function under general assumptions and present important instances in which the proximity operator can be computed explicitly or via straightforward numerical operations. These results constitute central building blocks in the design of proximal optimization algorithms. We showcase the versatility of the framework by designing novel proximal algorithms for state-of-the-art regression and variable selection schemes in high-dimensional statistics.

1. Introduction

Perspective functions appear, often implicitly, in various problems in areas as diverse as statistics, control, computer vision, mechanics, game theory, information theory, signal recovery, transportation theory, machine learning, disjunctive optimization, and physics (see the companion paper [7] for a detailed account). In the setting of a real Hilbert space $\mathcal{G}$, the most useful form of a perspective function, first investigated in Euclidean spaces in [24], is the following.

Definition 1.1. Let $\varphi\colon \mathcal{G} \to \,]-\infty, +\infty]$ be a proper lower semicontinuous convex function and let $\operatorname{rec}\varphi$ be its recession function. The perspective of φ is

* Corresponding author.
E-mail addresses: plc@math.ncsu.edu (P.L. Combettes), cmueller@simonsfoundation.org (C.L. Müller).

http://dx.doi.org/10.1016/j.jmaa.2016.12.021
0022-247X/

$$\widetilde{\varphi}\colon \mathbb{R}\times\mathcal{G}\to\,]-\infty,+\infty] : (\eta,y)\mapsto\begin{cases}\eta\varphi(y/\eta), & \text{if } \eta>0;\\ (\operatorname{rec}\varphi)(y), & \text{if } \eta=0;\\ +\infty, & \text{if } \eta<0.\end{cases} \tag{1.1}$$

Many scientific problems result in minimization problems that involve perspective functions. In statistics, a prominent instance is the modeling of data via "maximum likelihood-type" estimation (or M-estimation) with a so-called concomitant parameter [17]. In this context, φ is a likelihood function, η takes the role of the concomitant parameter, e.g., an unknown scale or location of the assumed parametric distribution, and y comprises unknown regression coefficients. The statistical problem is then to simultaneously estimate the concomitant variable and the regression vector from data via optimization. Another important example in statistics [15], signal recovery [5], and physics [16] is the Fisher information of a function $x\colon \mathbb{R}^N \to\,]0,+\infty[$, namely

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)}dt, \tag{1.2}$$

which hinges on the perspective function of the squared Euclidean norm (see [7] for further discussion).

In the literature, problems involving perspective functions are typically solved with a wide range of ad hoc methods. Despite the ubiquity of perspective functions, no systematic structuring framework has been available to approach these problems. The goal of this paper is to fill this gap by showing that they are amenable to solution by proximal methods, which offer a broad array of splitting algorithms to solve complex nonsmooth problems with attractive convergence guarantees [1,8,11,14]. The central element in the successful implementation of a proximal algorithm is the ability to compute the proximity operator of the functions present in the optimization problem. We therefore propose a systematic investigation of proximity operators for perspective functions and show that the proximal framework can efficiently solve perspective-function based problems, unveiling in particular new applications in high-dimensional statistics.

In Section 2, we introduce basic concepts from convex analysis and review essential properties of perspective function. We then study the proximity operator of perspective functions in Section 3. We establish a characterization of the proximity operator and then provide examples of computation for concrete instances. Section 4 presents new applications of perspective functions in high-dimensional statistics and demonstrates the flexibility and potency of the proposed framework to both model and solve complex problems in statistical data analysis.

2. Notation and background

2.1. Notation and elements of convex analysis

Throughout, $\mathcal{H}$, $\mathcal{G}$, and $\mathcal{K}$ are real Hilbert spaces and $\mathcal{H}\oplus\mathcal{G}$ denotes their Hilbert direct sum. The symbol $\|\cdot\|$ denotes the norm of a Hilbert space and $\langle\cdot\mid\cdot\rangle$ the associated scalar product. The closed ball with center $x\in\mathcal{K}$ and radius $\rho\in\,]0,+\infty[$ is denoted by $B(x;\rho)$.

A function $f\colon\mathcal{K}\to\,]-\infty,+\infty]$ is proper if $\operatorname{dom} f = \{x\in\mathcal{K} \mid f(x)<+\infty\}\neq\varnothing$, coercive if $\lim_{\|x\|\to+\infty} f(x)=+\infty$, and supercoercive if $\lim_{\|x\|\to+\infty} f(x)/\|x\|=+\infty$. Denote by $\Gamma_0(\mathcal{K})$ the class of proper lower semicontinuous convex functions from $\mathcal{K}$ to $]-\infty,+\infty]$, and let $f\in\Gamma_0(\mathcal{K})$. The conjugate of f is the function

$$f^*\colon\mathcal{K}\to[-\infty,+\infty] : u\mapsto\left(\sup_{x\in\mathcal{K}}\langle x\mid u\rangle - f(x)\right). \tag{2.1}$$

It also belongs to $\Gamma_0(\mathcal{K})$ and $f^{**}=f$. The subdifferential of f is the set-valued operator