Global Supervised Descent Method

Xuehan Xiong, Fernando De la Torre
The Robotics Institute, Carnegie Mellon University.

Mathematical optimization plays a fundamental role in solving many problems in computer vision (*e.g.*, camera calibration, image alignment, structure from motion). It is generally accepted that second order descent methods are the most robust, fast, and reliable approaches for nonlinear optimization of a general smooth function. However, in the context of computer vision, second order descent methods have two main drawbacks: 1) the function might not be analytically differentiable and numerical approximations are impractical, and 2) the Hessian may be large and not positive definite. Recently, Supervised Descent Method (SDM), a method that learns the "weighted averaged gradients" in a supervised manner has been proposed to solve these issues. However, SDM is a local algorithm and it is likely to average conflicting gradient directions. This paper proposes Global SDM (GSDM), an extension of SDM that divides the search space into regions of similar gradient directions.

Fig. 1a illustrates the idea of SDM. During training, in each iteration SDM learns a single generic Descent Map (DM) from the optimal optimization trajectories (indicated by the dotted lines). In testing, the same DM is used for driving an unseen sample to $\mathbf{x}_*$ (the labeled ground-truth). DM exists under two mild conditions proved in [3]. For simple functions with a unique minimum, such conditions normally hold. However, in many real applications the function might have several local minima in a relatively small neighborhood, for instance see Fig. 1b for an example. Standard SDM would average conflicting gradient directions resulting in undesirable performance. To overcome this issue, GSDM learns not one but a set of generic DMs (in this example, four), one for different domains (colored by different intensity of grays) of the objective function. Each domain contains only similar gradient directions and one separate DM is learned for each. At iteration k, $\mathbf{x}_k$ may step into any of the four regions and the corresponding DM is used to update.

Based on this intuition, this paper introduces and validates a new concept, *Domains of Homogeneous Descent (DHD)* and extends the theory of SDM to global optimization. In the paper, we prove that it is possible to find a partition of domain $\mathbf{x}$, such that there exists a generic DM for each subset. The subsets of this partition are defined as DHD. However, to guarantee the convergence of GSDM for a general function an exponential number of DMs are required to be learned. In practice, we rely on the following approximation. First, we apply dimension reduction techniques to the original data. Then, we create a partition where each subset occupies one of the hyper-octants in the reduced dimensional space, and a different DM is learned for each subset.

We develop a practical algorithm based on the above approximation to track faces from profile to profile. Our work differs from existing approaches in several ways. First, our approach do not pre-build any shape or appearance model and we directly optimize over landmark coordinates. This has been shown to provide superior performance for facial feature tracking [2]. Second, our method provides a mathematically sound manner to partition the parameter space for facial feature tracking. Existing approaches typically find heuristic partition of the head pose angles. Finally, our method is general and can be applied to other problems, such as extrinsic camera calibration. There is a lack of datasets for evaluation of face tracking from profile to profile as well as a standard protocol for evaluating tracking performance. To fill the void, we build two challenging datasets, Distracted Driver Face(DDF) and Naturalistic Driving Study(NDS), and propose a standard evaluation protocol for facial feature tracking. Both the evaluation protocol code and NDS dataset are made available for the research community at the following link: `http://humansensing.cs.cmu.edu/xxiong`. Our experiments show that GSDM is able to track more frames and provides more accurate landmark prediction than SDM on both datasets. Examples of tracking results can be watched from the link: `http://goo.gl/EGiUFV`.

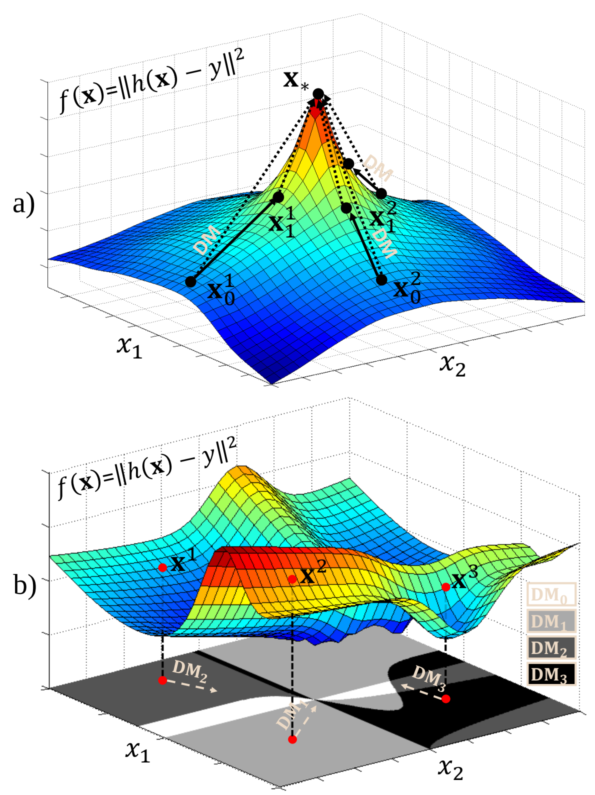

Figure 1: a) A single Descent Map (DM) is used in SDM for minimizing a simple function. b) An example of a more complex objective function. In order to use SDM, its domain has to be split into four regions (represented by different grays) and a separate DM is learned for each region.

Besides GSDM, we establish the connection between SDM and Imitation Learning (IL) [1]. SDM can be viewed as an algorithm for policy derivation and the DM from each step can be interpreted as a learned optimization policy in the context of IL. Since the ground truth solutions $\{\mathbf{x}_*^i\}$ are available throughout training, we can always receive the perfect feedbacks based on the state observation. SDM takes advantage of this fact by learning not one but a sequence of policies so the latter ones correct mistakes made from previous iterations. To the best of our knowledge, SDM is the first algorithm of IL applied to optimization. One of our ongoing research is to explore extensions of SDM as a policy derivation algorithm and use it to solve Robotics applications, such as inverse kinematics.

[1] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.

[2] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *Computer Vision and Pattern Recognition (CVPR)*, pages 532–539, 2013.

[3] Xuehan Xiong and Fernando De la Torre. Supervised descent method for solving nonlinear least squares problems in computer vision. *arXiv preprint arXiv:1405.0601*, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.