

PROBLEM DECOMPOSITION IN NONCONVEX OPTIMIZATION

Terry Rockafellar

University of Washington, Seattle

Flatiron Institute, New York

October 15, 2018

A Basic Format for Thinking About Decomposition

Problem in variables $x_j \in \mathbb{R}^{n_j}$

$$\text{minimize } \sum_{j=1}^J f_j(x_j) \quad \text{subject to } \sum_{j=1}^J F_j(x_j) \in K$$

Ingredients: for this presentation

functions $f_j : \mathbb{R}^{n_j} \rightarrow (-\infty, \infty]$, lower semicontinuous,
mappings $F_j : \mathbb{R}^{n_j} \rightarrow \mathbb{R}^m$, just $\mathcal{C}^1$ or $\mathcal{C}^2$,
cone $K \subset \mathbb{R}^m$ closed and convex

Challenges

- solve this by a scheme which breaks computations down into subproblems in the separate indices $j = 1, \dots, J$
- manage this, at least locally, even in the nonconvex case

Extensions: replace the cone constraint by a “coupling term”

Classical Lagrangian Approach

Lagrangian function: in terms of $Y = K^*$ = the **polar** cone

$$\begin{aligned} L(x_1, \dots, x_J, y) &= \sum_{j=1}^J f_j(x_j) + y \cdot \sum_{j=1}^J F_j(x_j) \text{ where } y \in Y \\ &= \sum_{j=1}^J L_j(x_j, y) \text{ for } L_j(x_j, y) = f_j(x_j) + y \cdot F_j(x_j) \end{aligned}$$

Convex case: exploit this separability via the dual problem,

$$\text{maximize } \sum_{j=1}^J g_j(y) \text{ over } y \in Y, \quad g_j(y) = \min_{x_j} L_j(x_j, y)$$

Augmented Lagrangian function: with parameter $r > 0$

$$\begin{aligned} L_r(x_1, \dots, x_J, y) &= \sup_{y' \in Y} \left\{ L(x_1, \dots, x_J, y') - \frac{1}{2r} \|y' - y\|^2 \right\} \\ &= L(x_1, \dots, x_J, y) + \frac{r}{2} \left\| \sum_{j=1}^J F_j(x_j) \right\|^2 - \frac{1}{2r} d_Y^2(y + r \sum_{j=1}^J F_j(x_j)) \end{aligned}$$

Augmented second-order sufficient condition for local optimality

$(\bar{x}_1, \dots, \bar{x}_J)$ and $\bar{y}$ satisfy the first-order optimality condition and, for r sufficiently large, L_r is convex-concave locally around them

Nonconvex case: convexity maybe gained, but separability lost

Reformulation to Help Liberate Separability

Expansion Lemma:

$$\sum_{j=1}^J F_j(x_j) \in K \iff \begin{cases} \exists u_j \in \mathbf{R}^m \text{ with } \sum_{j=1}^J u_j = 0 \\ \text{and } F_j(x_j) + u_j \in K \text{ for all } j \end{cases}$$

Expanded problem (equivalent) in variables x_j and u_j

$$\text{minimize } \sum_{j=1}^J f_j(x_j) \text{ subject to } F_j(x_j) + u_j \in K, \sum_{j=1}^J u_j = 0$$

First-order optimality condition: for $\bar{x}_j, \bar{u}_j$, to give a local min
 $\exists \bar{y} \in Y = K^*$ such that, for $j = 1, \dots, J$, the subproblem

$$\text{minimize } f_j(x_j) \text{ subject to } F_j(x_j) + \bar{u}_j \in K$$

exhibits first-order optimality at $\bar{x}_j$ with multiplier vector $\bar{y}$

- Status:**
- necessary under a constraint qualification
 - in the convex case, always sufficient

Progressive Decomposition Algorithm

Parameters: $s > 0$, $r > e \geq 0$ (where e “elicits convexity”)

Current elements (in iteration k): y^k , x_j^k , u_j^k with $\sum_{j=1}^J u_j^k = 0$

Optimization step: for $j = 1, \dots, J$, determine

$$(x_j^{k+1}, \hat{u}_j^k) \in (\text{local?}) \operatorname{argmin}_{x_j, u_j} \left\{ f_j^k(x_j, u_j) \mid F_j(x_j) + u_j \in K \right\}$$

$$\text{for } f_j^k(x_j, u_j) = f_j(x_j) + \frac{s}{2} \|x_j - x_j^k\|^2 + \frac{r}{2} \|u_j - u_j^k\|^2 - y^k \cdot u_j$$

Updating step: partner x_j^{k+1} with $u_j^{k+1} = u_j^k - \hat{u}_j^k$, where

$$\hat{u}^k = \frac{1}{J} \sum_{j=1}^J \hat{u}_j^k, \text{ and set } y^{k+1} = y^k - (r - e) \hat{u}^k$$

Behavior in the convex case, where e can be 0

The subproblems have $(x_j^{k+1}, \hat{u}_j^k) = \operatorname{argmin}$ (global, unique), and the sequences converge to globally optimal $\bar{x}_j$, $\bar{u}_j$, and multiplier $\bar{y}$

Behavior in the nonconvex case, with $\epsilon > 0$ sufficiently high

- The procedure needs to start near enough to elements that satisfy a new variational second-order sufficient condition
- Then $(x_j^{k+1}, \hat{u}_j^k) = \text{local argmin (locally unique)}$, and the sequences converge to locally optimal $\bar{x}_j$, $\bar{u}_j$, and multiplier $\bar{y}$

Challenges in the nonconvex case:

- How to know, in execution, whether ϵ is high enough?
- How to know if starting near enough to a solution?
- How to be sure of staying in the “good neighborhood”?
- How to choose the size of the parameters s and r , or adjust?

Challenges even for the convex case:

- What stopping criterion other than “exact minimization”?
- Is it possible to pass from parallel computation in the separate subproblems to some “asynchronous” version?

Connection Back to Augmented Lagrangians

Reconfiguring the optimization step: in minimizing

$$f_j^k(x_j, u_j) = f_j(x_j) + \frac{s}{2} \|x_j - x_j^k\|^2 + \frac{r}{2} \|u_j - u_j^k\|^2 - y^k \cdot u_j$$

in (x_j, u_j) subject to $F_j(x_j) + u_j \in K$ to get $(x_j^{k+1}, \hat{u}_j^k)$,

(1) first minimize in u_j , then (2) minimize the residual in x_j

Observation: the residual is $L_{j,r}^k(x_j, y^k) + \frac{s}{2} \|x_j - x_j^k\|^2$, where

$L_{j,r}^k(x_j, y) =$ augmented Lagrangian for the subproblem:

minimize $f_j(x_j)$ subject to $F_j(x_j) + u_j^k \in K$

Resulting form of the optimization step in iteration k

$$x_j^{k+1} = (\text{local}) \operatorname{argmin}_{x_j} \{ L_{j,r}^k(x_j, y^k) + \frac{s}{2} \|x_j - x_j^k\|^2 \},$$

$$\hat{u}_j^k = -\nabla_y L_{j,r}^k(x_j^{k+1}, y^k) \quad \text{augmented separability achieved!}$$

Resulting interpretation of elicitation:

e = level of r inducing the augmented second-order sufficient condition — as adapted to the extended problem formulation

Dealing With Nonconvex Nonsmoothness

Consider: $g : \mathbb{R}^n \rightarrow (-\infty, \infty]$ lsc, $\bar{x} \in \text{dom } g$

Local minimum?

1st-order necessary condition: $0 \in \partial g(\bar{x})$

2nd-order variational sufficient condition: **variational convexity**

“ ∂g is indistinguishable in graph around $(\bar{x}, 0)$ from $\partial \hat{g}$ for a convex function $\hat{g}$, and the values of g and $\hat{g}$ agree there”

strong version of this: $\hat{g}$ strongly convex

- corresponds to local max monotonicity of ∂g
- strong version closely related to tilt stability of g at $\bar{x}$
- tied to ability to work locally with the prox mapping for g

A Key Favorable Class of Functions

for working with variational convexity that isn't just local convexity

Strongly amenable functions: $g : \mathbf{R}^n \rightarrow (-\infty, \infty]$
“representable locally as $h \circ G$ for h convex, $G \in \mathcal{C}^2$ under a c.q.”

Example: $g = g_0 + \delta_C$ (locally) with

- $g_0 = \max$ of finitely many $\mathcal{C}^2$ functions
- C specified by finitely many $\mathcal{C}^2$ constraints

$\implies$ then “prox” can be executed by nonlinear programming

References

- [1] R.T. Rockafellar (2018) “Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity,” *Set-Valued and Variational Analysis*, accepted <http://dx.doi.org/10.1007/s1128-018-0496-1>
- [2] R.T. Rockafellar (2018) “Variational convexity and local monotonicity of subgradient mappings,” *Vietnam Journal of Mathematics*, accepted (special issue in honor of A. D. Ioffe).
- [3] R.T. Rockafellar, R. J-B Wets (1998) *Variational Analysis*, Springer Grundlehren der Mathematischen Wissenschaften 317.

downloads: sites.math.washington.edu/~rtr/mypage.html