Augmented Lagrangians + Decomposition in Convex and Nonconvex Programming

Terry Rockafellar
University of Washington, Seattle

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
October 15, 2018

An Optimization Model for Promoting Decomposition

Problem

minimize $\sum_{j=1}^{q} f_j(x_j) + g\left(\sum_{j=1}^{q} F_j(x_j)\right)$ over $(x_1, \ldots, x_q) \in S$

Ingredients: for this presentation

- mappings $F_j : \boldsymbol{R}^{n_j} \to \boldsymbol{R}^m$, just $\mathcal{C}^1$ or $\mathcal{C}^2$ for $j = 1, \ldots, q$,
- functions $f_j : \boldsymbol{R}^{n_j} \to (-\infty, \infty]$, just lsc for $j = 1, \ldots, q$,
- function $g : \boldsymbol{R}^m \to (-\infty, \infty]$, lsc, convex, pos. homogeneous
- subspace $S \subset \boldsymbol{R}^n = \boldsymbol{R}^{n_1} \times \cdots \times \boldsymbol{R}^{n_q}$ with complement $S^\perp$

Challenge

solve this by a scheme which breaks computations down into subproblems in separate indices j that bypass the S constraint

Territory Covered by this Formulation

minimize $\sum_{j=1}^{q} f_j(x_j) + g\Big(\sum_{j=1}^{q} F_j(x_j)\Big)$ over $(x_1,\ldots,x_q) \in S$

Specializations of the coupling term:

- $g(u) = \delta_K(u)$ for a closed convex cone K for a constraint
- $g(u) = ||u|| =$ some norm for regularization term

pos. homogeneity of g can be dropped with some adjustments

Specializations of the coupling space:

- S gives application-dependent linear relations among the x_j's
- $S = \big\{(x_1,\ldots,x_q) \,\big|\, x_1 = \cdots = x_q\big\}$, for the splitting version
- S taken to be all of $\boldsymbol{R}^n$ (thereby "dropping out"), $S^\perp = \{0\}$

Specializatins to convex optimization:

- f_j convex and $F_j = A_j$ affine
- f_j and F_j convex and g nondecreasing among others

Reformulation to Liberate Underlying Separability

Expansion Lemma

$$g\left(\sum_{j=1}^q F_j(x_j)\right) \le \alpha \quad \Longleftrightarrow \quad \exists\, u_j \in \boldsymbol{R}^m \text{ for } j = 1,\dots,q$$

$$\text{such that } \sum_{j=1}^q u_j = 0 \text{ and } \sum_{j=1}^q g\big(F_j(x_j) + u_j\big) \le \alpha$$

Extended coupling space: now in $\boldsymbol{R}^n \times [\boldsymbol{R}^m]^q$

$$\overline{S} = \left\{(x_1,\dots,x_q,u_1,\dots,u_q) \,\middle|\, (x_1,\dots,x_q) \in S,\ \sum_{j=1}^q u_j = 0\right\},$$

$$\overline{S}^\perp = \left\{(v_1,\dots,v_q,y_1,\dots,y_q) \,\middle|\, (v_1,\dots,v_q) \in S^\perp,\ y_1 = \cdots = y_q\right\}$$

Expanded problem (equivalent)

$$\min \sum_{j=1}^q \Big[\, f_j(x_j) + g\big(F_j(x_j) + u_j\big) \Big] \text{ over } (x_1,\dots,x_q,u_1,\dots,u_q) \in \overline{S}$$

$\longrightarrow$ **separability achieved in the objective:**

$$\varphi(x_1,\dots,x_q,u_1,\dots,u_q) = \varphi_1(x_1,u_1) + \cdots + \varphi_q(x_q,u_q)$$

Linkage Problems in Terms of Subgradients

Goal: minimize some lsc function f over some subspace S
to be applied later to minimizing φ on $\overline{S}$ as above

First-order condition for local optimality

$$\bar{w} \in S \ \text{ and } \ \exists\, \bar{z} \in \partial f(\bar{w}) \ \text{ such that } \ \bar{z} \in S^{\perp}$$

Regular subgradients: notation $\bar{z} \in \widehat{\partial} f(\bar{w})$

$$f(w) \geq f(\bar{w}) + \bar{z}\cdot(w - \bar{w}) + o(||w - \bar{w}||)$$

General subgradients: notation $\bar{z} \in \partial f(\bar{w})$

$$\exists\, z^{\nu} \to \bar{z} \ \text{ with } \ z^{\nu} \in \widehat{\partial} f(w^{\nu}),\ w^{\nu} \to \bar{w},\ f(w^{\nu}) \to f(\bar{w})$$

Convex case: general = regular = convex subgradients
Smooth case: general = regular = classical gradients

Linkage problem — for given f and S

$$\text{find a pair } (\bar{w}, \bar{z}) \in [\operatorname{gph} \partial f] \cap [S \times S^{\perp}]$$

Second-order Sufficiency via Virtual Convexity

Key observation: in terms of e = "elicitation" parameter ≥ 0,
$d_S(w)$ = distance of w from the subspace S

minimizing f on S $\quad\longleftrightarrow\quad$ minimizing $f_e = f + \frac{e}{2}d_S^2$ on S

First-order optimality is thereby unaffected:

$\bar{z} \in \partial f(\bar{w}) \iff \bar{z} \in \partial f_e(\bar{w})$ when $\bar{w} \in S$ and $\bar{z} \in S^\perp$

Variational second-order sufficient condition: in addition,
for e high enough, f_e is variationally convex at $(\bar{w}, \bar{z})$, meaning

$\exists\, \varepsilon > 0$, open convex nbhd $W \times Z$ of $(\bar{w}, \bar{z})$, and lsc convex
$h \leq f_e$ on W such that $\mathrm{gph}\, \partial h$ coincides in $W \times Z$ with

$$\mathrm{gph}\, T_{e,\varepsilon} = \left\{ (w, z) \in \mathrm{gph}\, \partial f_e \,\middle|\, f_e(w) \leq f_e(\bar{w}) + \varepsilon \right\}$$

and, on that common set, furthermore $h(w) = f_e(w)$

Strong version: the function $h \leq f_e$ is <u>strongly</u> convex

Sufficiency in the Convex and Smooth Cases

Convex example

for convex f, the variational condition is superfluous

the first-order condition already guarantees <u>global</u> optimality

Smooth example: $f \in \mathcal{C}^2$ with gradient $\nabla f(\bar{w})$, hessian $\nabla^2 f(\bar{w})$

- the first-order condition reduces to:

  $\bar{w} \in S$, and the gradient $\bar{z} = \nabla f(\bar{w})$ is $\perp S$

- the second-order condition in strong form reduces to:

  $\nabla^2 f(\bar{w})$ is positive definite relative to S

$\longrightarrow$ these are the standard sufficient conditions for a local min

Progressive Decoupling of Linkages (Rock. 2018)

for determining $(\bar{w}, \bar{z}) \in [\mathrm{gph}\,\partial f] \cap [S \times S^{\perp}]$

Algorithm with parameters $r > e \geq 0$, generating $\{(w^k, z^k)\}_{k=1}^{\infty}$

In iteration k, having $w^k \in S$ and $z^k \in S^{\perp}$, get

$$\widehat{w}^k = (\text{local?})\,\mathrm{argmin}_w \left\{ f(w) - z^k \cdot w + \tfrac{r}{2}||w - w^k||^2 \right\}$$

Update by $\quad w^{k+1} = \mathrm{proj}_S\, \widehat{w}^k, \qquad z^{k+1} = z^k - (r-e)[\widehat{w}^k - w^{k+1}]$

Convergence Theorem

Convex case: converges globally from any initial (w^0, z^0)

General case: if $(\bar{w}, \bar{z})$ satisfies the sufficient condition at elicitation level e, then $\exists$ nbhd $W \times Z$ of $(\bar{w}, \bar{z})$ such that, if $(w^0, z^0) \in W \times Z$, the generated sequence stays in $W \times Z$ with $\widehat{w}^k =$ unique local minimizer on W, and it converges to to some solution $(\widetilde{w}, \widetilde{z})$ such that $\widetilde{w} \in$ argmin of f on $W \cap S$

Underpinnings of the Progressive Decoupling Algorithm

- exploits properties of max monotonicity of set-valued mappings
- derives from the proximal point algorithm of Rock. (1976)
- extends the partial inverse method of Spingarn (1983)
- extends the proximal point localization of Pennanen (2002)

Criterion for local max monotonicity — Rock. (2018)

The variational sufficiency condition $\Longrightarrow$ the mapping
$T_{e,\varepsilon}$ having its graph $= \{(w,z) \in \operatorname{gph} \partial f_e \mid f_e(w) \leq f_e(\bar{w}) + \varepsilon\}$
is locally **max monotone** around $(\bar{w}, \bar{z})$, and moreover is **equivalent** to that when $\bar{z}$ is a **regular** subgradient of f at $\bar{w}$

$\Longrightarrow$ **the proximal point algorithm can operate locally**
as long as the initial (w^0, z^0) is near enough to $(\bar{w}, \bar{z})$

Application to the Expanded Programming Model

minimize $\varphi(x_1, \ldots, x_q, u_1, \ldots, u_q) = \sum_{j=1}^{q} \varphi_j(x_j, u_j)$ over $\overline{S}$

$$\text{where} \;\; \varphi_j(x_j, u_j) \;=\; f_j(x_j) + g\big(F_j(x_j) + u_j\big)$$

$$\overline{S} = \big\{(x_1, \ldots, x_q, u_1, \ldots, u_q) \,\big|\, (x_1, \ldots, x_q) \in S, \;\; \textstyle\sum_{j=1}^{q} u_j = 0\big\},$$

$$\overline{S}^{\perp} = \big\{(v_1, \ldots, v_q, y_1, \ldots, y_q) \,\big|\, (v_1, \ldots, v_q) \in S^{\perp}, \; y_1 = \cdots = y_q\big\}$$

Algorithm elements in this specialization:

$w^k = (x_1^k, \ldots, x_q^k, u_1^k, \ldots, u_q^k)$ for $(x_1^k, \ldots, x_q^k) \in S$, $\sum_{j=1}^{q} u_j^k = 0$,

$z^k = (v_1^k, \ldots, v_q^k, y^k, \ldots, y^k)$ for $(v_1^k, \ldots, v_q^k) \in S^{\perp}$

Decomposition property from liberated separability

The step in which the algorithm determines $\widehat{w}^k$ breakes down for $j = 1 \ldots, q$ to calculating: $(\widehat{x}_j^k, \widehat{u}_j^k) =$ (local?)argmin of

$$\varphi_j^k(x_j, u_j) = \varphi_j(x_j, u_j) - (v_j^k, y^k)\cdot(x_j, u_j) + \tfrac{r}{2}||(x_j, u_j) - (x_j^k, u_j^k)||^2$$

Resulting Procedure — Full Form

Algorithm (with parameters $r > e \geq 0$)

In iteration k, having $(x_1^k, \ldots, x_q^k) \in S$ and $(v_1^k, \ldots, v_q^k) \in S^\perp$ along with y^k and $(u_1^k, \ldots, u_q^k)$ such that $\sum_{j=1}^q u_j^k = 0$,

determine $(\widehat{x}_j^k, \widehat{u}_j^k)$ for $j = 1, \ldots, q$ as the (local?) minimizer of
$f_j(x_j) + g(F_j(x_j) + u_j) - v_j^k \cdot x_j - y^k \cdot u_j + \frac{r}{2}||x_j - x_j^k||^2 + \frac{r}{2}||u_j - u_j^k||^2$

Then let $\widehat{u}^k = \frac{1}{q} \sum_{j=1}^q \widehat{u}_j^k$ and update by

$$(x_1^{k+1}, \ldots, x_q^{k+1}) = \text{proj}_S(\widehat{x}_j^k, \ldots, \widehat{x}_j^k), \qquad u_j^{k+1} = u_j^k - \widehat{u}^k$$
$$v_j^{k+1} = v_j^k - (r - e)[\widehat{x}_j^k - x_j^{k+1}], \qquad y^{k+1} = y^k - (r - e)\widehat{u}^k$$

Convergence: global in the convex case, and moreover local in the nonconvex case as long as the algorithm starts near enough to a solution where the second-order variational sufficiency condition is satisfied at level e of the elicitation parameter

Bringing in Augmented Lagrangians

Consider auxiliary subproblems:

minimize $f_j(x_j) + g(F_j^k(x_j))$ in x_j where $F_j^k(x_j) = F_j(x_j) + u_j^k$

Dualization: g is lsc convex pos.homog., so its conjugate is $g^* = \delta_Y$ (indicator) for some closed convex set $Y \subset R^m$

Examples: $g = \delta_K$ for cone K yields Y = polar cone K^*
$g = ||\cdot||_p$ yields Y = unit ball for dual norm $||\cdot||_q$

Lagrangians: $L_j^k(x_j, y) = f_j(x_j) + y \cdot F_j^k(x_j) - \delta_Y(y)$

Augmented Lagrangians (with parameter $r > 0$):

$$L_{j,r}^k(x_j, y) = f_j(x_j) + y \cdot F_j^k(x_j) + \tfrac{r}{2}||F_j^k(x_j)||^2 - \tfrac{1}{2r} d_Y^2\big(y + rF_j^k(x_j)\big)$$
$$= f_j(x_j) + \min_{u_j} \left\{ g(F_j(x_j) + u_j) - y \cdot u_j + \tfrac{r}{2}||u_j - u_j^k||^2 \right\}$$

Key observation: this min arises in the algorithm for $y = y^k$
$\longrightarrow$ and then $\widehat{u}_j^k$, the argmin, equals $-\nabla_{y_j} L_{j,r}^k(x_j, y_j^k)$

Resulting Procedure with Augmented Lagrangians

Decomposition algorithm in condensed form

From $(x_1^k, \ldots, x_q^k) \in S$, $(v_1^k, \ldots, v_q^k) \in S^\perp$, $\sum_{j=1}^q u_j^k = 0$, y^k, get

$$\widehat{x}_j^k = (\text{local})\, \text{argmin}_{x_j} \left\{ L_{j,r}^k(x_j, y^k) - v_j^k \cdot x_j + \frac{r}{2} ||x_j - x_j^k||^2 \right\}$$

and update by $(x_1^{k+1}, \ldots, x_q^{k+1}) = \text{proj}_S(\widehat{x}_j^k, \ldots, \widehat{x}_j^k)$,

$v_j^{k+1} = v_j^k - (r-e)[\widehat{x}_j^k - x_j^{k+1}]$, $\widehat{u}_j^k = -\nabla_y L_{j,r}^k(x_j^{k+1}, y^k)$,

$\widehat{u}^k = \frac{1}{q} \sum_{j=1}^q \widehat{u}_j^k$, $u_j^{k+1} = u_j^k - \widehat{u}^k$, $y^{k+1} = y^k - (r-e)\widehat{u}^k$

Note: a convenient formula for the gradient is often available

Connection with the new second-order local optimality criterion

The variational sufficiency condition holds for a solution with elements $\bar{x}_j$, $\bar{v}_j$, $\bar{u}_j$, $\bar{y}$, with respect to an elicitation level e if and only if there are neighborhoods $X_j \times Y_j$ of $(\bar{x}_j, \bar{y})$ such that the iterations have $L_{j,r}^k(x_j, y)$ convex-concave on $X_j \times Y_j$

References

[1] R.T. Rockafellar (2018) "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," accepted for publication.

[2] R.T. Rockafellar (2018) "Variational convexity and local monotonicity of subgradient mappings," accepted for publication.

[3] R.T. Rockafellar (2018) "Variational second-order sufficiency, generalized augmented Lagrangians and local duality in optimization," soon to be available.

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