Convergence Analysis of BALM

João Xavier*, Alessio Del Bue†, Lourdes Agapito, Marco Paladini‡

October 7, 2011

Abstract

In this report, we provide a convergence analysis for the algorithm BALM introduced in "Bilinear modelling via Augmented Lagrange Multipliers (BALM)", by Alessio del Bue, João Xavier, Lourdes Agapito and Marco Paladini, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

*João Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal. Corresponding author: `jxavier@isr.ist.utl.pt`

†Alessio Del Bue is with the Istituto Italiano di Tecnologia (IIT), Genova, Italy.

‡Lourdes Agapito and Marco Paladini are with the School of Electronic Engineering and Computer Science (EECS) of Queen Mary University of London, UK.

1 Introduction

In the paper "Bilinear modelling via Augmented Lagrange Multipliers (BALM)", by Alessio del Bue, João Xavier, Lourdes Agapito and Marco Paladini, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, we introduced the BALM algorithm. For lack of space, a convergence analysis was not included. Here, we provide a convergence analysis of BALM (see section 3), which was labeled Algorithm 1 in the aforementioned paper. As an intermediate result, we also study the convergence of Algorithm 2 in the BALM paper (see section 2).

2 Block coordinate descent

We consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X \end{array} \tag{1}$$

where X is the Cartesian product

$$X = X_1 \times \cdots \times X_{m+1},$$

with $X_i \subset \mathbb{R}^{n_i}$ for $i = 1, \ldots, m+1$. The vector x is partitioned as $x = (x_1, \ldots, x_{m+1})$ and $n = n_1 + \cdots + n_{m+1}$. Let $x = (x_1, \ldots, x_{m+1}) \in X$. For each $i = 1, \ldots, m+1$, we define $f_i(\cdot; x) : \mathbb{R}^{n_i} \to \mathbb{R}$,

$$f_i(\xi; x) = f\left(x_1, \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_{m+1}\right).$$

We say that $x \in X$ is regular for f if for each $i = 1, \ldots, m$ the function $f_i(\cdot; x)$ has an unique minimizer over X_i. Note the asymmetric treatment: the function $f_{m+1}(\cdot; x)$ is not involved.

We make the following assumptions:

A) for each $i = 1, \ldots, m$, there holds $X_i = \mathbb{R}^{n_i}$, and X_{m+1} denotes an embedded submanifold of $\mathbb{R}^{n_{m+1}}$ (e.g., see [2] for this basic notion of smooth manifold theory);

B) the function f is continuously differentiable over an open set containing X. Moreover, $f_i(\cdot; x)$ is a convex function for all $i = 1, \ldots, m$ and $x \in X$.

First-order necessary conditions. Given these assumptions, if $x = (x_1, \ldots, x_{m+1}) \in X$ is a local minimizer of f then it satisfies the following first-order necessary conditions:

$$\nabla f_i\left(x_i; x\right) = 0, \quad i = 1, \ldots, m, \tag{2}$$

$$\nabla f_{m+1}\left(x_{m+1}; x\right) \in N_{x_{m+1}} X_{m+1}, \tag{3}$$

where $N_{x_{m+1}} X_{m+1}$ denotes the normal space to the manifold X_{m+1} at the point $x_{m+1} \in X_{m+1}$, i.e., the linear subspace of $\mathbb{R}^{n_{m+1}}$ which is orthogonal to $T_{x_{m+1}} X_{m+1}$, the tangent space to

X_{m+1} at the point x_{m+1}. Let us further clarify (3). Commonly, the embedded submanifold X_{m+1} is represented as a level set of a submersion, i.e., $X_{m+1} = \{\xi \in \mathbb{R}^{n_{m+1}} : h(\xi) = 0\}$ where $h : \mathbb{R}^{n_{m+1}} \to \mathbb{R}^p$ $(p < n_{m+1})$, $h(\xi) = (h_1(\xi), \ldots, h_p(\xi))$ denotes a submersion: a smooth (infinitely differentiable) map with full rank Jacobian everywhere, that is, $\operatorname{rank} \nabla h(\xi) = p$ for all ξ, where $\nabla h(\xi)$ denotes the $n_{m+1} \times p$ matrix

$$\nabla h(\xi) = \begin{bmatrix} \nabla h_1(\xi) & \cdots & \nabla h_p(\xi) \end{bmatrix}$$

(equivalently, the constraint gradients $\nabla h_1(\xi), \ldots, \nabla h_p(\xi)$ are linearly independent). The manifold X_{m+1} has dimension $n_{m+1} - p$ and the tangent space to X_{m+1} at the point x_{m+1} is given by

$$T_{x_{m+1}} X_{m+1} = \text{Kernel}\left(\nabla h(x_{m+1})^\top\right),$$

see proposition lemma 8.15 in [2]. It follows that the normal space is

$$N_{x_{m+1}} X_{m+1} = \text{Range}\left(\nabla h(x_{m+1})\right) = \left\{ \sum_{j=1}^{p} \nabla h_j(x_{m+1}) \lambda_j \ : \lambda_j \in \mathbb{R} \text{ for } j = 1, \ldots, p \right\}.$$

Therefore, (3) translates into $\nabla f_{m+1}(x_{m+1}; x) = \sum_{j=1}^{p} \nabla h_j(x_{m+1}) \lambda_j$, for some $\lambda_j \in \mathbb{R}$, which is the usual form of expressing the necessary first-order conditions for the problem

$$\begin{array}{ll} \text{minimize} & f_{m+1}(\xi; x) \\ \text{subject to} & h(\xi) = 0, \end{array} \tag{4}$$

e.g., see proposition 3.1.1 in [1]. Note that x_{m+1} solves (4).

We need to keep the generic notation (3) because in the nonrigid SfM and articulated SfM applications, the associated manifolds are not (globally) represented as level sets of submersions.

An important property. The assumed convexity of $f_i(\cdot; x)$ for all $i = 1, \ldots, m$ and $x \in X$ yields the following property, which will turn out important for the convergence analysis of the block coordinate descent method: let $i \in \{1, \ldots, m\}$ and let $\overline{\xi} \in X_i$ be a minimizer of $f_i(\cdot; x)$ over X_i; then, $f_i(\cdot; x)$ is monotonically nonincreasing in the interval from x_i to $\overline{\xi}$, i.e.,

$$f_i\left((1 - t_2) x_i + t_2 \overline{\xi}; x\right) \leq f_i\left((1 - t_1) x_i + t_1 \overline{\xi}; x\right) \quad \text{for all} \quad 0 \leq t_1 \leq t_2 \leq 1. \tag{5}$$

This is readily established as follows. Let $\phi(t) = f_i\left((1 - t) x_i + t\overline{\xi}; x\right) - f_i\left(\overline{\xi}; x\right)$. Note that $\phi(t)$ is a convex function of t. Moreover, $\phi(t) \geq 0$ for all t. Now, let $0 \leq t_1 < t_2 \leq 1$ (the case $t_1 = t_2$ is trivial in (5)). Write $t_2 = (1 - \alpha) t_1 + \alpha 1$ where $\alpha = (t_2 - t_1)/(1 - t_1)$ belongs to $[0, 1]$. The convexity and the nonnegativity of ϕ yield

$$\phi(t_2) \leq (1 - \alpha)\phi(t_1) + \alpha\phi(1) = (1 - \alpha)\phi(t_1) \leq \phi(t_1),$$

which establishes (5).

Block coordinate descent method. The block coordinate descent method (BCD) operates as follows. Given the current iterate

$$x^k = \left(x_1^k, \ldots, x_{m+1}^k\right) \in X,$$

the next iterate $x^{k+1} = \left(x_1^{k+1}, \ldots, x_{m+1}^{k+1}\right) \in X$ is generated as

$$x_i^{k+1} \in \arg\min_{\xi \in X_i} f\left(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, \xi, x_{i+1}^{k}, \ldots, x_{m+1}^{k}\right), \quad i = 1, \ldots, m+1. \tag{6}$$

Theorem 1. *(Convergence of block coordinate descent) Let assumptions A,B hold for problem* (1)*. Let* $\{x^k\}$ *be a sequence obeying the recursion* (6)*. Then, every limit point of* $\{x^k\}$ *which is regular for* f *satisfies the first-order conditions* (2)-(3)*.*

Before delving into the proof, an important remark. Note that the theorem says that, in particular, if the sequence $\{x^k\}$ generated by the BCD method happens to converge, then its limit (assumed regular for f) will satisfy the first-order necessary conditions for a local minimizer of the optimization problem (1). The theorem is not powerful enough to assert that $\{x^k\}$ will converge nor that, if it does, the limit is a local minimizer for (1). The conclusion of the theorem is what can be typically guaranteed for a BCD method, without further assumptions, e.g., compare theorem 1 with proposition 2.7.1 in [1]. However, in practice, when $\{x^k\}$ converges, its limit is usually indeed a local minimizer for (1).

Proof of theorem 1: We just follow the reasoning of the proof of proposition 2.7.1 in [1] (the proof in the errata available at `http://www.athenasc.com/nonlinbook.html`) and add some new lines to handle the manifold X_{m+1}.

Let $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_{m+1})$ be a point of $\{x^k\}$ which is regular for f. Let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ which converges to $\overline{x}$. By replicating the steps of the proof in [1] we can conclude that

$$\nabla f_i\left(\overline{x}_i; \overline{x}\right) = 0, \quad i = 1, \ldots, m, \tag{7}$$

and that $\{x_i^{k_j+1}\}$ converges to $\overline{x}_i$ as $j \to \infty$, for each $i = 1, \ldots, m$. We omit here the details of this replication, since they are a straightforward adaptation of the aforementioned existing proof. We just remark that the reasoning in [1] can be replicated here because property (5) holds and that $\overline{x}$ is regular for f.

Equation (7) shows that $\overline{x}$ satisfies (2) in the first-order necessary conditions (2)-(3). We now show that it also satisfies (3). We have

$$f\left(x^{k_{j+1}}\right) \le f\left(x^{k_j+1}\right) \le f\left(x_1^{k_j+1}, \ldots, x_m^{k_j+1}, \xi\right), \text{ for all } \xi \in X_{m+1}.$$

Taking the limit $j \to \infty$ gives

$$f(\overline{x}) \le f(\overline{x}_1, \ldots, \overline{x}_m, \xi), \quad \text{for all } \xi \in X_{m+1},$$

that is, $\overline{x}_{m+1}$ solves the optimization problem

$$\begin{array}{ll} \text{minimize} & f_{m+1}(\xi; \overline{x}) \\ \text{subject to} & \xi \in X_{m+1}. \end{array} \tag{8}$$

Now, since X_{m+1} is an embedded submanifold of $\mathbb{R}^{n_{m+1}}$ it can be represented near $\overline{x}_{m+1}$ as a level set of a submersion, see proposition 8.12 in [2]. More precisely, there exist an open set $U \subset \mathbb{R}^{n_{m+1}}$ containing $\overline{x}_{m+1}$, and a submersion $h : U \to \mathbb{R}^p$ ($p < n_{m+1}$), $h(x) = (h_1(x), \ldots, h_p(x))$, such that $X_{m+1} \cap U = \{\xi \in U : h(\xi) = 0\}$. Moreover, the tangent space to the manifold X_{m+1} at any point $\xi \in X_{m+1} \cap U$ is given by $T_\xi X_{m+1} = \text{Kernel}\left(\nabla h(\xi)^\top\right)$ and the corresponding normal space is $N_\xi X_{m+1} = \text{Range}\left(\nabla h(\xi)\right) = \{\nabla h(\xi)\lambda : \lambda \in \mathbb{R}^p\}$. Since $\overline{x}_{m+1}$ solves (8), it also solves the restricted version

$$\begin{array}{ll} \text{minimize} & f_{m+1}(\xi; \overline{x}) \\ \text{subject to} & h(\xi) = 0, \xi \in U. \end{array} \tag{9}$$

Thus it satisfies the first-order necessary conditions of proposition 3.1.1 in [1], i.e.,

$$\nabla f_{m+1}(\overline{x}_{m+1}; \overline{x}) = \sum_j \nabla h_j(\overline{x}_{m+1})\lambda_j$$

for some $\lambda_j \in \mathbb{R}$, which is equivalent to (3) ■

Application to BALM. We now show that Algorithm 2 (iterative Gauss Seidel) in the BALM paper fits into the framework of theorem 1. More precisely, we shall consider the following algorithm 2.1.

Algorithm 2.1 corresponds to Algorithm 2 in the BALM paper, modulo these irrelevant modifications: 1) the order of the minimizations is changed - we update the manifold variable at the end (in order to match the framework of theorem 1); 2) two minimizations, one for S and another for M, are done, instead of a single joint minimization for (S, M) as in Algorithm 2; as explained in the BALM paper, this is another possible variation of Algorithm 2, which, in fact, was used in our numerical experiments; 3) we let the algorithm run forever (we removed the upper bound $L_{\max}$ on the number of iterations) in order to study the asymptotic behavior of the infinite sequence $\left\{\left(z^{[l]}, S^{[l]}, M^{[l]}, N^{[l]}\right)\right\}$ generated by the iterative Gauss Seidel iterations.

In terms of theorem 1, we have $m + 1 = 4$ block variables, i.e., $x = (z, S, M, N)$ with $z \in X_1 = \mathbb{R}^q$ (q is the number of missing entries), $S \in X_2 = \mathbb{R}^{n\times r}$, $M \in X_3 = \mathbb{R}^{r\times m}$ and $N = (N_1, \ldots, N_f) \in X_4 = \mathcal{M}^f$, where $\mathcal{M}$ denotes the particular manifold at hand and $\mathcal{M}^f$ denotes the Cartesian product $\mathcal{M} \times \cdots \times \mathcal{M}$ (f factors). Also, the cost function is $f(x) = f(z, S, L, M) = L_{\sigma^{(k)}}\left(z, S, M, N; R^{(k)}\right)$, where

$$L_\sigma(z, S, M, N) = \|Y(z) - SM\|^2 - \mathsf{tr}\left(R^\top(M - N)\right) + \frac{\sigma}{2}\|M - N\|^2 .$$

We now check the assumptions A) and B) needed in the theorem. Assumption B) is trivially satisfied. Indeed, f is obviously a smooth function. Moreover, when f is viewed as a function of z alone (resp. S and M), it is convex, i.e., $f_1(\cdot; x)$ (resp. $f_2(\cdot; x)$ and $f_3(\cdot; x)$) is convex: in fact, it is a simple quadratic for all $x = (z, S, M, N)$. Thus assumption B) holds. Interestingly, note that $f_1(\cdot; x)$ and $f_3(\cdot; x)$ are strictly convex for all x and $f_2(\cdot; x)$ is also strictly convex for those $x = (z, S, M, N)$ in which M has full row rank. Since rank deficient

Algorithm 2.1 (algorithm 2 in the BALM paper)

set $l = 0$

set $z^{[0]} = z^{(k)}$, $S^{[0]} = S^{(k)}$ and $M^{[0]} = M^{(k)}$

loop

solve

$$\begin{aligned} z^{[l+1]} &= \\ &= \operatorname{argmin} \quad L_{\sigma^{(k)}} \left(z, S^{[l]}, M^{[l]}, N^{[l]}; R^{(k)}\right) \end{aligned} \tag{10}$$

solve

$$\begin{aligned} S^{[l+1]} &= \\ &= \operatorname{argmin} \quad L_{\sigma^{(k)}} \left(z^{[l+1]}, S, M^{[l]}, N^{[l]}; R^{(k)}\right) \end{aligned} \tag{11}$$

solve

$$\begin{aligned} M^{[l+1]} &= \\ &= \operatorname{argmin} \quad L_{\sigma^{(k)}} \left(z^{[l+1]}, S^{[l+1]}, M, N^{[l]}; R^{(k)}\right) \end{aligned} \tag{12}$$

solve

$$\begin{aligned} N^{[l+1]} &= \\ &= \begin{array}{ll} \operatorname{argmin} & L_{\sigma^{(k)}} \left(z^{[l+1]}, S^{[l+1]}, M^{[l+1]}, N; R^{(k)}\right) \\ \text{subject to} & N_i \in \mathcal{M}, \quad i = 1, \ldots, f, \end{array} \end{aligned} \tag{13}$$

update $l \leftarrow l + 1$

end loop

matrices $M \in \mathbb{R}^{r\times m}$ constitute a zero measure set in $\mathbb{R}^{r\times m}$, we see that all points x (except for a zero measure set) are regular for f. In that sense, the assumption of regularity of limit points mentioned in theorem 1 is very mild.

We now address assumption A). Since the Cartesian product of embedded submanifolds is an embedded submanifold it suffices to show that $\mathcal{M}$ is an embedded submanifold of $\mathbb{R}^{r\times p}$. We do it for the three cases of interest in the BALM paper:

1. **Nonrigid structure from motion.** We consider the set

$$\mathcal{M} = \{t \otimes Q \,:\, t = (t_1, \ldots, t_d) \in \mathbb{R}^d, Q^\top Q = I_2\} - \{0\}. \tag{14}$$

This is not exactly the set $\mathcal{M}$ defined in equation (14) of the BALM paper. We removed one point (the origin), which is irrelevant in practice (but would raise a topological difficulty in our study). We now prove that $\mathcal{M}$ in (14) is an embedded submanifold of $\mathbb{R}^{3d\times 2} - \{0\}$. We assume that the reader is familiar with basic concepts of smooth manifold theory, e.g., at the level of [2]. The set $\mathcal{M}$ in (14) can be equivalently represented as

$$\mathcal{M} = \{\rho\, u \otimes Q \,:\, \rho > 0, u \in \mathsf{S}^{d-1}, Q^\top Q = I_2\}, \tag{15}$$

where $\mathsf{S}^{d-1} = \{u \in \mathbb{R}^d \,:\, \|u\| = 1\}$ denotes the unit-sphere in $\mathbb{R}^d$. For reasons that will be apparent later, let us introduce the smooth manifold $P = \mathsf{S}^{d-1} \times \mathsf{O}(3, 2)$ where $\mathsf{O}(3, 2) = \{Q \in \mathbb{R}^{3\times 2} \,:\, Q^\top Q = I_2\}$ denotes the Stiefel manifold of 3×2 matrices. Furthermore, consider the (discrete) Lie group $G = \{\pm 1\}$ acting on P as follows: $(u, Q) \overset{g}{\mapsto} (gu, gQ)$ for all $g \in G$ and $(u, Q) \in P$. It is straightforward to check that this action is smooth, free and proper, see [2]. Thus, invoking theorem 9.16 in [2], the quotient space P/G can be turned into a smooth manifold such that the canonical map $\pi \,:\, P \to P/G$ is a smooth, surjective submersion. Note that the quotient space is nothing more than the set of equivalence classes of P induced by the equivalence relation: $(u_1, Q_1) \sim (u_2, Q_2)$ if $(u_1, Q_1) = \pm(u_2, Q_2)$. We denote by $[(u, Q)](= \pi(u, Q))$ a typical element of P/G. Consider the smooth manifold $A = \mathbb{R}_{++} \times (P/G)$ where $\mathbb{R}_{++}$ is the set of strictly positive reals. The set $\mathcal{M}$ is (15) can be written as $\mathcal{M} = F(A)$ where $F \,:\, A \to \mathbb{R}^{3d\times 2} - \{0\}$ is defined as $F(\rho, [(u, Q)]) = \rho u \otimes Q$. According to theorem 8.3 in [2], $\mathcal{M}$ is an embedded submanifold of $\mathbb{R}^{3d\times 2} - \{0\}$ if F is a smooth embedding.

We now focus on proving that F is a smooth embedding. We show that we can invoke proposition 7.4 (b) in [2]. First, it is trivial to check that F is injective (that's why the quotient space was introduced). Second, note that $\widehat{F} \,:\, \mathbb{R}_{++} \times P \to \mathbb{R}^{3d\times 2} - \{0\}$, $\widehat{F}(\rho, (u, Q)) = \rho u \otimes Q$ is clearly smooth and the equality $\widehat{F} = F \circ \lambda$ holds, where $\lambda \,:\, \mathbb{R}_{++} \times P \to A$, $\lambda(\rho, (u, Q)) = (\rho, \pi(u, Q))$ is a surjective smooth submersion. Thus, we may invoke proposition 7.17 in [2] to conclude that F is smooth. Third, F is an immersion. This can be established as follows. We begin by showing that F has constant rank. Consider the group $G = \mathbb{R}_{++} \times \mathsf{O}(d) \times \mathsf{O}(3)$ with group multiplication $(s_1, V_1, W_1) \cdot (s_2, V_2, W_2) = (s_1 s_2, V_1 V_2, W_1 W_2)$ and where $\mathsf{O}(n) = \{V \in \mathbb{R}^{n\times n} \,:\, V^\top V = I_n\}$ denotes the set of $n \times n$ orthogonal matrices. It is clear that G is a smooth

Lie group assuming the canonical smooth manifold structures of the Lie groups $\mathbb{R}_{++}$ and $\mathsf{O}(n)$. Define an action of G on A as $(\rho, [(u, Q)]) \overset{(s,V,W)}{\mapsto} (s\rho, [(Vu, WQ)])$. It is straightforward to check that this action is smooth and transitive. Also, define an action of G on $\mathbb{R}^{3d\times 2} - \{0\}$ as $Z \overset{(s,V,W)}{\mapsto} s(V \otimes W)Z$. Clearly, this action is smooth. Finally, note that $F : A \to \mathbb{R}^{3d\times 2}$ is equivariant with respect to those two actions. It follows from theorem 9.7 in [2] that F has constant rank. Now, since F is also injective, theorem 7.15 (b) in [2] asserts that F is an immersion. Fourth and finally, we show that F is a proper map. Let K be a compact subset of $\mathbb{R}^{3d\times 2} - \{0\}$. We must show that $F^{-1}(K) = \{(\rho, [(u, Q)]) \in A : F(\rho, [(u, Q)])\} \in K$ is compact. Equivalently, we must show that any sequence in $F^{-1}(K)$ admits a convergent subsequence. Let $\{(\rho^k, [(u^k, Q^k)])\}$ be a sequence in $F^{-1}(K)$. Since $\{Z^k\}$, where $Z^k = F(\rho^k, [(u^k, Q^k)])$, is a sequence in the compact set K, it admits a convergent subsequence. Thus, by restricting to a subsequence if necessary, we may assume that $\{Z^k\}$ converges, say to $\overline{Z} \in \mathbb{R}^{3d\times 2} - \{0\}$ as $k \to \infty$. Now, it is easily seen that $\rho^k = \left\|Z^k\right\| / \sqrt{2}$. Thus, $\{\rho^k\}$ converges to $\left\|\overline{Z}\right\| / \sqrt{2}$ as $k \to \infty$. Finally, note that P/G is compact since it the image of the compact manifold $P = \mathsf{S}^{d-1} \times \mathsf{O}(3, 2)$ by the continuous map π. Thus, passing to a subsequence if necessary, we may further assume that $\{[(u^k, Q^k)]\}$ is convergent. We conclude that $(\rho^k, [(u^k, Q^k)])$ is convergent.

We proved that $\mathcal{M}$ in (15) is an embedded submanifold $\mathbb{R}^{3d\times 2} - \{0\}$. Since the later is itself an embedded submanifold of $\mathbb{R}^{3d\times 2}$ (because it is an open subset of $\mathbb{R}^{3d\times 2}$), we conclude that $\mathcal{M}$ in (15) is an embedded submanifold of $\mathbb{R}^{3d\times 2}$.

2. **Articulated structure from motion.** Consider the set

$$\mathcal{M} = \{(u, A, B) \in \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} : uu^\top + AA^\top = uu^\top + BB^\top = I_2\} - \mathcal{Z} \quad (16)$$

where $\mathcal{Z} = \{(u, A, B) : \det A = 0\}$. This is the set $\mathcal{M}$ defined in equation (22) of the BALM paper, except that we removed a (closed) zero measure set $\mathcal{Z}$, which is irrelevant in practice. We will prove that $\mathcal{M}$ in (16) is an embedded submanifold of $U = \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} - \mathcal{Z}$ by applying corollary 8.10 in [2]. Introduce the map $\Phi : U \to \mathsf{S}(2) \times \mathsf{S}(2)$,

$$\Phi(u, A, B) = \left(uu^\top + AA^\top, uu^\top + BB^\top\right),$$

where $\mathsf{S}(n) = \{X \in \mathbb{R}^{n\times n} : X = X^\top\}$ is the linear subspace of $n \times n$ symmetric matrices, viewed here as a smooth manifold. Note that $\mathcal{M} = \Phi^{-1}(I_2, I_2)$. We proceed to show that (I_2, I_2) is a regular value of Φ. It is clear that Φ is smooth. Let $(u, A, B) \in \Phi^{-1}(I_2, I_2)$. We must show that the derivative map of Φ, evaluated at (u, A, B), is a surjective linear map. It is straightforward to check that this derivative map corresponds to the linear map $\Phi_* : \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} \to \mathsf{S}(2) \times \mathsf{S}(2)$,

$$\Phi_*(\delta_u, \Delta_A, \Delta_B) = \left(\delta_u u^\top + u\delta_u^\top + \Delta_A A^\top + A\Delta_A^\top, \delta_u u^\top + u\delta_u^\top + \Delta_B B^\top + B\Delta_B^\top\right).$$

Now, let $(X, Y) \in \mathsf{S}(2) \times \mathsf{S}(2)$. We show that there exist $(\delta_u, \Delta_A, \Delta_B)$ such that $\Phi_*(\delta_u, \Delta_A, \Delta_B) = (X, Y)$. We let $\delta_u = 0$. Consider the equation

$$\Delta_A A^\top + A \Delta_A^\top = X. \tag{17}$$

Since A is nonsingular, we can change variables as $\Omega_A = \Delta_A A^\top$, and (17) turns into

$$\Omega_A + \Omega_A^\top = X, \tag{18}$$

where

$$X = \begin{bmatrix} X_a & X_b \\ X_b & X_c \end{bmatrix} \in \mathsf{S}(2)$$

is given and

$$\Omega_A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

is to be found. But, clearly, the choice $\alpha = X_a/2$, $\beta = X_b$, $\gamma = 0$ and $\delta = X_c/2$ solves (18), and $\Delta_A = \Omega_A A^{-\top}$ solves (17). Now, consider the equation

$$\Delta_B B^\top + B \Delta_B^\top = Y$$

to be solved for Δ_B. We could duplicate the previous reasoning to show that there exists a solution, if B were nonsingular. But, indeed, B is nonsingular. Note that $(u, A, B) \in \Phi^{-1}(I_2, I_, 2)$. Thus, the two equations $uu^\top + AA^\top = I_2$, $uu^T + BB^\top = I_2$ hold. Subtracting one from the other yields $AA^\top = BB^\top$. Since A is nonsingular, B is also nonsingular.

In sum, we have shown that $\mathcal{M}$ is an embedded submanifold of U. But, since U is an open subset of $\mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2}$ it is an embedded submanifold of $\mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2}$. Thus, $\mathcal{M}$ is an embedded submanifold of $\mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2}$.

3. **Photometric stereo.** The relevant set here is

$$\mathcal{M} = \{\rho(1, z) \,:\, \rho \in \mathbb{R}, z \in \mathbb{R}^3, z^\top z = 1\} - \{0\}. \tag{19}$$

Compared to $\mathcal{M}$ given in equation (25) of the BALM paper, we have just removed one point (the origin), which is irrelevant in practice. We now prove that $\mathcal{M}$ in (19) is an embedded submanifold of $\mathbb{R}^4 - \{0\}$. Note that $\mathcal{M}$ can be represented as $\mathcal{M} = F^{-1}(0)$ where $F : \mathbb{R}^4 - \{0\} \to \mathbb{R}$ is given by $F(\xi_1, \xi_2, \xi_3, \xi_4) = \xi_1^2 - (\xi_2^2 + \xi_3^2 + \xi_4^2)$. Now, it is straightforward to check that F is a smooth submersion. Thus, according to corollary 8.9 in [2], $\mathcal{M}$ is an embedded submanifold of $\mathbb{R}^4 - \{0\}$. Since the later manifold is itself an embedded submanifold of $\mathbb{R}^4$, we conclude that $\mathcal{M}$ in (19) is an embedded submanifold of $\mathbb{R}^4$.

3 BALM

Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x_1, x_2) \\ \text{subject to} & x_2 \in X \end{array} \tag{20}$$

where $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$ and $X \subset \mathbb{R}^{n_2}$ denotes an embedded submanifold of $\mathbb{R}^{n_2}$. Moreover, $f : \mathbb{R}^{n_1+n_2} \to \mathbb{R}$ is assumed to be continuously differentiable. Problem (20) is equivalent to

$$\begin{array}{ll} \text{minimize} & f(x_1, x_2) \\ \text{subject to} & x_2 = x_3 \\ & x_3 \in X. \end{array} \tag{21}$$

Also, it is straightforward to see that $(\overline{x}_1, \overline{x}_2)$ is a local minimizer for problem (20) if and only if $(\overline{x}_1, \overline{x}_2, \overline{x}_2)$ is a local minimizer for problem (21). We now write the first-order necessary conditions that must hold at a minimizer $(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ for problem (21). Since $\overline{x}_3 \in X$ and X is an embedded submanifold of $\mathbb{R}^{n_2}$, proposition 8.12 in [2] asserts the existence of an open set $U \subset \mathbb{R}^{n_2}$ containing $\overline{x}_3$ such that $X \cap U = \{\xi \in U : h(\xi) = 0\}$ for some smooth (infinitely differentiable) map $h : U \to \mathbb{R}^p$, $h(x) = (h_1(x), \ldots, h_p(x))$, where p denotes the codimension of the submanifold X (i.e., $\dim X = n_2 - p$). Additionally, h is a submersion, that is, $\operatorname{rank} \nabla h(\xi) = p$ for all $\xi \in U$, where $\nabla h(\xi) = \begin{bmatrix} \nabla h_1(\xi) & \cdots & \nabla h_p(\xi) \end{bmatrix}$. Thus, if $(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ is a local minimizer for problem (21) then it is also a local minimizer for the problem

$$\begin{array}{ll} \text{minimize} & f(x_1, x_2) \\ \text{subject to} & x_2 = x_3 \\ & h(x_3) = 0 \\ & x_3 \in U. \end{array} \tag{22}$$

Now, using proposition 3.1.1 in [1], the necessary first-order conditions for (22) are given by

$$\nabla_{x_1} f(\overline{x}_1, \overline{x}_2) = 0 \tag{23}$$
$$\nabla_{x_2} f(\overline{x}_1, \overline{x}_2) + \overline{\alpha} = 0 \tag{24}$$
$$-\overline{\alpha} + \nabla h(\overline{x}_3)\overline{\beta} = 0 \tag{25}$$
$$\overline{x}_2 = \overline{x}_3, h(\overline{x}_3) = 0, \tag{26}$$

for some Lagrange multiplier $(\overline{\alpha}, \overline{\beta}) \in \mathbb{R}^{n_2} \times \mathbb{R}^p$.

Augmented Lagrangian approach. We consider an augmented Lagrangian optimization method to solve (21), see algorithm 3.1. The augmented Lagrangian is given by

$$L_\sigma(x_1, x_2, x_3; \lambda) = f(x_1, x_2) - \lambda^\top (x_2 - x_3) + \frac{\sigma}{2} \|x_2 - x_3\|^2 .$$

Thus, only the first equality constraint in (21) has been dualized, the manifold constraint is present in each subproblem (27) of the augmented Lagrangian method.

Algorithm 3.1 (augmented Lagrangian Method to solve (21))

set $k = 0$ and $\epsilon_{\mathsf{best}} = +\infty$
initialize $\sigma^0 > 0$, $\lambda^0 \in \mathbb{R}^{n_2}$, $\gamma > 1$ and $0 < \eta < 1$
initialize $x_1^0 \in \mathbb{R}^{n_1}$ and $x_2^0 \in \mathbb{R}^{n_2}$
loop
    solve

$$\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}\right) = \begin{array}{ll} \arg\min & L_{\sigma^k}(x_1, x_2, x_3; \lambda^k) \\ \text{subject to} & x_3 \in X \end{array} \tag{27}$$

    compute $\epsilon = \left\| x_2^{k+1} - x_3^{k+1} \right\|^2$
    if $\epsilon < \eta\, \epsilon_{\mathsf{best}}$
        $\lambda^{k+1} = \lambda^k - \sigma^k \left(x_2^{k+1} - x_3^{k+1}\right)$
        $\sigma^{k+1} = \sigma^k$
        $\epsilon_{\mathsf{best}} = \epsilon$
    else
        $\lambda^{k+1} = \lambda^k$
        $\sigma^{k+1} = \gamma \sigma^k$
    endif
    update $k \leftarrow k + 1$
end loop

Theorem 2. *(Convergence of augmented Lagrangian method) Suppose $\{\lambda^k\}$ is bounded. Then, every limit point $(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ of the sequence $\{(x_1^k, x_2^k, x_3^k)\}$ satisfies the first-order conditions* (23)-(26).

Both the assumption and the conclusion match the spirit of proposition 4.2.2 in [1], but note that theorem 2 refers to the sequence $\{(x_1^k, x_2^k, x_3^k)\}$ generated by algorithm 3.1.

Proof of theorem 2: We essentially follow the reasoning of the proof of proposition 4.2.2 in [1] with a minor adaptation to account for algorithm 3.1. Let $\{(x_1^{k_j}, x_2^{k_j}, x_3^{k_j})\}$ be a subsequence of (x_1^k, x_2^k, x_3^k) which converges to $(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ as $j \to \infty$. Let the embedded submanifold $X \subset \mathbb{R}^{n_2}$ be represented near $\overline{x}_3$ as a level set, i.e., $X \cap U = \{\xi \in U : h(\xi) = (h_1(\xi), \ldots, h_p(\xi)) = 0\}$ for some open set $U \subset \mathbb{R}^{n_2}$ containing $\overline{x}_3$ and smooth map $h : U \to \mathbb{R}^p$ with $\operatorname{rank} \nabla h(\xi) = p$ for all $\xi \in U$ (existence of such U and h is guaranteed by proposition 8.12 in [2]). Since $x_3^{k_j} \to \overline{x}_3$ we may assume without loss of generality that the whole sequence $\{x_3^{k_j}\}$ is contained in U. Also, since $\{\lambda^{k_j - 1}\}$ is a bounded sequence, by passing to a subsequence if necessary, we may assume that it converges, say to $\overline{\lambda}$, as $j \to \infty$.

Now, according to (27), $(x_1^{k_j}, x_2^{k_j}, x_3^{k_j})$ is a local minimizer for

$$\begin{array}{ll} \text{minimize} & L_{\sigma^{k_j - 1}}(x_1, x_2, x_3; \lambda^{k_j - 1}) \\ \text{subject to} & h(x_3) = 0 \\ & x_3 \in U. \end{array}$$

Thus, proposition 3.1.1 in [1] asserts the existence of a Lagrange multiplier $\beta^j \in \mathbb{R}^p$ such that

$$\begin{aligned}
\nabla_{x_1} f(x_1^{k_j}, x_2^{k_j}) &= 0 & (28)\\
\nabla_{x_2} f(x_1^{k_j}, x_2^{k_j}) - \lambda^{k_j-1} + \sigma^{k_j-1}(x_2^{k_j} - x_3^{k_j}) &= 0 & (29)\\
\lambda^{k_j-1} + \sigma^{k_j-1}(x_3^{k_j} - x_2^{k_j}) + \nabla h(x_3^{k_j})\beta^j &= 0 & (30)
\end{aligned}$$

hold. Taking the limit in (28) gives

$$\nabla_{x_1} f(\overline{x}_1, \overline{x}_2) = 0. \tag{31}$$

Define $\gamma^j = \sigma^{k_j-1}(x_2^{k_j} - x_3^{k_j})$. Thus, equation (29) corresponds to

$$\nabla_{x_2} f(x_1^{k_j}, x_2^{k_j}) - \lambda^{k_j-1} + \gamma^j = 0,$$

and it follows that $\gamma^j \to \overline{\gamma}$ as $j \to \infty$, where $\overline{\gamma} = \overline{\lambda} - \nabla_{x_2} f(\overline{x}_1, \overline{x}_2)$. Taking the limit in (29) gives

$$\nabla_{x_2} f(\overline{x}_1, \overline{x}_2) + \overline{\alpha} = 0, \tag{32}$$

where we defined $\overline{\alpha} = -\overline{\lambda} + \overline{\gamma}$. Equation (30) corresponds to

$$\lambda^{k_j-1} - \gamma^j + \nabla h(x_3^{k_j})\beta^j = 0,$$

which, since $\nabla h(x_3^{k_j})$ has full column rank, implies that

$$\beta^j = \left(\nabla h(x_3^{k_j})^\top \nabla h(x_3^{k_j})\right)^{-1} \nabla h(x_3^{k_j})^\top \left(\gamma^j - \lambda^{k_j-1}\right).$$

Thus, $\beta^j \to \overline{\beta} = \left(\nabla h(\overline{x}_3)^\top \nabla h(\overline{x}_3)\right)^{-1} \nabla h(\overline{x}_3)^\top \left(\overline{\gamma} - \overline{\lambda}\right)$ and, taking limits in (30), gives

$$-\overline{\alpha} + \nabla h(\overline{x}_3)\overline{\beta} = 0. \tag{33}$$

Note that (31)-(33) establish (23)-(25). Since we also have $h(\overline{x}_3) = 0$, it only remains to show that $\overline{x}_2 = \overline{x}_3$ in order to prove that the whole set of necessary conditions (23)-(26) hold. We split the analysis in two cases:

1. $\sigma^{k_j-1} \to \infty$: since $\gamma^j = \sigma^{k_j-1}(x_2^{k_j} - x_3^{k_j})$ is convergent, we conclude that $x_2^{k_j} - x_3^{k_j} \to 0$. Thus, $\overline{x}_2 = \overline{x}_3$;

2. σ^{k_j} is upper bounded: by the nature of the algorithm 3.1 this means that, after some finite k_0, $\sigma^k = \sigma^{k_0}$ for all $k \geq k_0$ which implies that

$$\left\| x_2^{k+1} - x_3^{k+1} \right\|^2 < \eta \left\| x_2^k - x_3^k \right\|^2$$

for $k > k_0$. Since $\eta < 1$ we conclude that $\overline{x}_2 = \overline{x}_3$ ■

Application to BALM. Note that algorithm 3.1 corresponds to Algorithm 1 in the BALM paper, after one makes the identification $x_1 = (z, S)$, $x_2 = M$, $x_3 = N$ and $X = \mathcal{M} \times \mathcal{M} \times \cdots \times \mathcal{M}$ (f times). As the proof of theorem 2 shows, it suffices that (x_1^k, x_2^k, x_3^k) satisfy the first-order necessary conditions for subproblem (27) for the conclusion of theorem 2 to hold. That is, it is not required that (x_1^k, x_2^k, x_3^k) is a global minimizer for subproblem (27). This is in line with theorem 1, which (under the appropriate assumptions) guarantees that (x_1^k, x_2^k, x_3^k) satisfies those first-order necessary conditions.

References

[1] D. Bertsekas, *Nonlinear Programming*, 2nd ed, Athena Scientific (see errata for proposition 2.7.1 at `http://www.athenasc.com/nonlinbook.html`).

[2] J. Lee, *Introduction to Smooth Manifolds*, Springer.