Some Lifting Notes

The lifting method discussed in [8] applies to fairly general nonconvex quadratic problems that have the form of minimizing a quadratic objective subject to multiple quadratic constraints,

$$\begin{aligned} \min_x \; & x^T Q_0 x + b_0^T x \\ & x^T Q_j x + b_j^T x + c_j = 0, \quad j = 1, ..., m \; . \end{aligned} \tag{P}$$

This problem is equivalent to

$$\begin{aligned} \min_{x,X} \; & \mathrm{tr}\,(Q_0 X) + b_0^T x \\ & \mathrm{tr}\,(Q_j X) + b_j^T x + c_j = 0, \quad j = 1, ..., m \\ & X = xx^T . \end{aligned}$$

The convex semidefinite program obtained by relaxing the constraint $X = xx^T$ to $X \succeq xx^T$ is given by

$$\begin{aligned} \min_{x,X} \; & \mathrm{tr}\,(Q_0 X) + b_0^T x \\ & \mathrm{tr}\,(Q_j X) + b_j^T x + c_j = 0, \quad j = 1, ..., m \\ & \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \; . \end{aligned} \tag{SDP}$$

It is shown in [8] that (SDP) can be derived by computing a bidual of (P). They construct a Lagrangian

$$L(x, p) = x^T Q_0 x + b_0^T x + \sum_j p_j (x^T Q_j x + b_j^T x + c_j)$$

and dual problem

$$\sup_p \inf_x L(x, p) \; , \tag{D}$$

which as it turns out is also equivalent to a semidefinite program. (SDP) can be obtained by constructing a dual of the dual problem (D).

This leads to bounds on the optimal values of these problems,

$$\mathrm{val}\,(\mathrm{P}) \geq \mathrm{val}\,(\mathrm{SDP}) \geq \mathrm{val}\,(\mathrm{D}) \; .$$

In fact $\mathrm{val}\,(\mathrm{SDP}) = \mathrm{val}\,(\mathrm{D})$ in the usual case that strong duality holds there. Altogether, the gap between (P) and (SDP) is bounded by the duality gap between (P) and (D).

If a rank one minimizer of the lifted convex relaxation (SDP) can be found, it should be a global minimum of (P). To study whether we can find a rank one minimizer, we might try to adapt recent arguments in [1, 5, 4, 6] that for similar problems construct dual certificates that guarantee the existence of a unique rank one minimizer. As in [1], in which lifting is applied to a blind deconvolution problem, it may also be helpful to represent the unknowns in lower dimensional subspaces.

Unfortunately, we can't solve (SDP) directly. It is far too large for the kinds of problems we are interested in. As a more tractable alternative, we will replace X in (SDP) by LL^T for a rank r matrix L and solve

$$\begin{aligned} \min_{L} \ & \operatorname{tr}(Q_0(LL^T)_{11}) + b_0^T(LL^T)_{12} \\ & \operatorname{tr}(Q_j(LL^T)_{11}) + b_j^T(LL^T)_{12} + c_j = 0 \\ & (LL^T)_{22} = 1 \ , \end{aligned} \tag{FACr}$$

where $(LL^T)_{11}$ represents X and $(LL^T)_{12}$ represents x.

This low rank factorization strategy has been used to effectively minimize convex semidefinite relaxations for SVD-free matrix completion [2] and interferometric inversion [7] for example. Its use is justified by a result of Burer and Monteiro [3] which says that local minima of (FACr) coincide with local minima of (SDP) under an additional rank constraint, namely

$$\begin{aligned} \min_{Y} \ & \operatorname{tr}(Q_0 Y_{11}) + b_0^T Y_{12} \\ & \operatorname{tr}(Q_j Y_{11}) + b_j^T Y_{12} + c_j = 0, \quad j = 1, ..., m \\ & Y \succeq 0 \\ & Y_{22} = 1 \\ & \operatorname{rank} Y \leq r \ , \end{aligned} \tag{SDPr}$$

where Y represents $\begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix}$, Y_{11} represents X and Y_{12} represents x.

Although going from (SDPr) to (FACr) doesn't introduce additional local minima, there is still a question of how local minima of (SDPr) relate to those of (SDP) and (P). We need to take a closer look at [3] regarding classification of local minima of (SDPr) to understand precisely when they are global minima of (SDP), but roughly speaking it looks like if r is bigger than the rank of the optimal solution of (SDP) then local minima of (FACr) are global minima of (SDP). So if it can be shown that (SDP) has a unique minimizer that is rank one, then as long as $r \geq 2$, a local minimum of (FACr) should be a rank one global minimizer of (SDP) and hence of (P).

The success of partially lifting the problem (P) to (FACr) seems to hinge on two things:

1. Does (SDP) have a unique rank one minimizer?
2. Can we find a rank one minimizer of (FACr)?

Rank two relaxations have been successfully used to avoid local minima when solving related semidefinite programs with rank one constraints, for example in [1] and [7]. It may be interesting to apply a similar strategy to the penalty method proposed by van Leeuwen and Herrmann for full waveform inversion in [9]. Formally, the problem has the form

$$\min_{u,m} \sum_{s,v} \frac{1}{2} \|Pu_{sv} - d_{sv}\|^2 + \frac{\lambda^2}{2} \|Lu_{sv} + w_v^2 \operatorname{diag}(u_{sv}) m - q_{sv}\|^2 \ ,$$

which can be written as (P). (In practice, there is an issue with the boundary conditions, but I think there is a workaround.)

Abstractly, the penalty method objective can be written as a bilinear least squares problem

$$\min_{x,y} \sum_i \frac{c_i}{2} \|A_i \operatorname{diag}(x) y + B_i x + C_i y + D_i\|^2 \ , \tag{BLS}$$

which can be written in the form of (P) if we replace $\operatorname{diag}(x)y$ by z and add the constraints $z = \operatorname{diag}(xy^T)$. The SDP relaxation of (BLS) is

$$\begin{aligned} \min_Y \ \operatorname{tr} & \left(Y \sum_i \begin{bmatrix} B_i \\ C_i \\ A_i \\ D_i \end{bmatrix} \begin{bmatrix} B_i^T & C_i^T & A_i^T & D_i^T \end{bmatrix} \right) \\ & Y \succeq 0 \\ & Y_{34} = \operatorname{diag}(Y_{12}) \\ & Y_{44} = 1 \ , \end{aligned} \tag{SDPBLS}$$

where Y represents $\begin{bmatrix} xx^T & xy^T & xz^T & x \\ yx^T & yy^T & yz^T & y \\ zx^T & zy^T & zz^T & z \\ x^T & y^T & z^T & 1 \end{bmatrix}$.

As a possible alternative with fewer unknowns, we can substitute $z = \operatorname{diag}(xy^T)$ back into the objective and write (SDPBLS) as

$$\begin{aligned} \min_Y \ & \sum_i \frac{c_i}{2} \|A_i \operatorname{diag}(Y_{12}) + B_i Y_{13} + C_i Y_{23} + D_i\|^2 \\ & Y \succeq 0 \\ & Y_{33} = 1 \ , \end{aligned} \tag{QBLS}$$

where Y represents $\begin{bmatrix} xx^T & xy^T & x \\ yx^T & yy^T & y \\ x^T & y^T & 1 \end{bmatrix}$. Since the objective is now quadratic, (QBLS) is no longer written as a standard semidefinite program, but it is still equivalent to (SDPBLS). The factorized problem analogous to (FACr) that we propose to solve replaces Y in either (SDPBLS) or (QBLS) with LL^T for a rank 2 matrix L.

Preliminary experiments on small toy examples indicate that the lifting idea can work for problems of the form (BLS). For example,

$$\min_{x,y} \frac{1}{2}(xy-1)^2 + \frac{\beta}{2}(y+x-8)^2 + \frac{\gamma+\mu}{2}x^2 + \mu y^2$$

has distinct minima roughly near $(0,8)$ and $(8,0)$ for certain positive parameter choices, and for small γ the minimum near $(0,8)$ has a slightly smaller value. If we apply a gradient method directly to the nonconvex problem starting near $(8,0)$, we get stuck in the nearby local minimum.

But if we apply a gradient method to the lifted form of (QBLS) analogous to (FACr) with a 3 by 2 matrix L, then we move away from that local minimum and converge instead to the global minimum. The path taken by the iterates is shown in Figure 1, and the values of the original and lifted objectives are shown in Figure 2. The computed L is very close to being rank one, in which case LL^T represents $\begin{bmatrix} xx^T & xy^T & x \\ yx^T & yy^T & y \\ x^T & y^T & 1 \end{bmatrix}$.

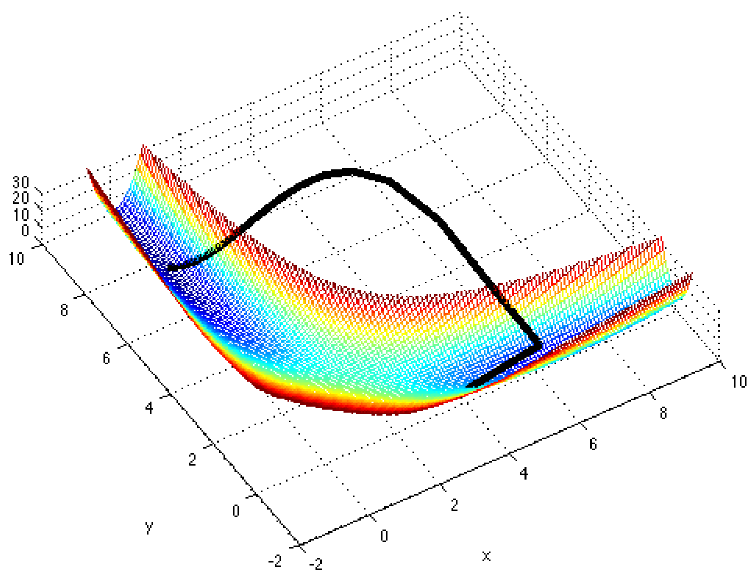

Figure 1: Gradient descent iterates of the rank 2 lifting are mapped to the original nonconvex surface. They move away from the local minimum on the right towards the global minimum on the left.

A possible difficulty is that the lifted problem appears to be very poorly conditioned when there are local minima with values very close to the value of the global minimum. Even for a simple two dimensional problem, a gradient method can take thousands of iterations. In order to be computationally tractable for more realistic examples, we probably need to use something like a Gauss Newton method.

References

[1] A. Ahmed, B. Recht, and J. Romberg, *Blind deconvolution using convex programming*, arXiv preprint arXiv:1211.5608, (2012), p. 40.

[2] A. Aravkin, R. Kumar, H. Mansour, and B. Recht, *A robust SVD-free approach to matrix completion, with applications to interpolation of large scale data*, (2013), pp. 1–20.

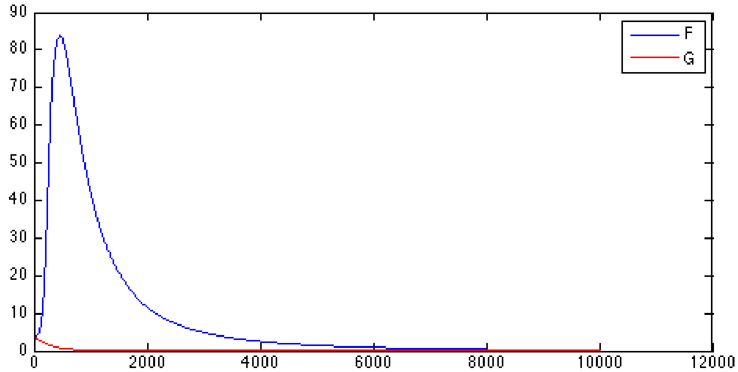

Figure 2: The lifted objective in L (red) decreases monotonically while the nonconvex objective in $x = L_{13}$ and $y = L_{23}$ (blue) first decreases as iterates move towards the nearby local minimum, then increases, and finally decreases to its global minimum.

[3] S. Burer and R. D. Monteiro, *Local Minima and Convergence in Low-Rank Semidefinite Programming*, Mathematical Programming, 103 (2004), pp. 427–444.

[4] E. J. Candes and Y. Plan, *Matrix Completion With Noise*, Proceedings of the IEEE, 98 (2010), pp. 925–936.

[5] E. J. Candès, T. Strohmer, and V. Voroninski, *PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming*, Communications on Pure and Applied Mathematics, 66 (2013), pp. 1241–1274.

[6] L. Demanet and P. Hand, *Stable optimizationless recovery from phaseless linear measurements*, arXiv preprint arXiv:1208.1803, (2012), pp. 1–20.

[7] L. Demanet and V. Jugnon, *Convex recovery from interferometric measurements*, arXiv preprint arXiv:1307.6864, (2013), pp. 1–19.

[8] C. Lemaréchal and F. Oustry, *Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization*, (1999).

[9] T. van Leeuwen and F. J. Herrmann, *Mitigating local minima in full-waveform inversion by expanding the search space*, Geophysical Journal International, 195 (2013), pp. 661–667.