Supplementary Materials: Multi-sample aCGH Data Analysis via Total Variation and Spectral Regularization

Xiaowei Zhou*, Can Yang, Xiang Wan, Hongyu Zhao and Weichuan Yu

1 More experiments

1.1 Parameter selection

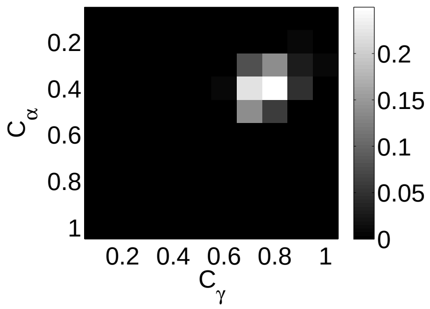
Here, we give a test to the parameter selection procedure. As described in Section 2.4 in the paper, the parameters are selected upon random sampling to divide the data into a training set and a test set. Each time of sampling may select different $\hat{\alpha}$ and $\hat{\gamma}$. To test the stability of parameter selection, we run 100 times of random sampling for a given data set and plot the distribution of selected parameters in Figure 1(a). As we can see, there is only one centroid with a small variance. Since the formulation of our method is continuous and convex, the small perturbation of parameters will not cause large changes of final results. Hence, sampling once should be stable enough to choose appropriate parameters.

Next, we illustrate the effectiveness of the proposed parameter tuning method. We generate various data sets with different shared percentages and apply our algorithm on them. Figure 1(b) shows the dependency between the selected $\hat{\alpha}$ and the shared percentage. As the shared percentage increases, $\hat{\alpha}$ also gets larger, which gives more emphasis to the nuclear norm penalty in Equation 5. This shows that the automatically selected parameters are adaptive to the underlying structure of input data. This flexibility is important in real applications, since it is difficult to know the true property of data before processing it.

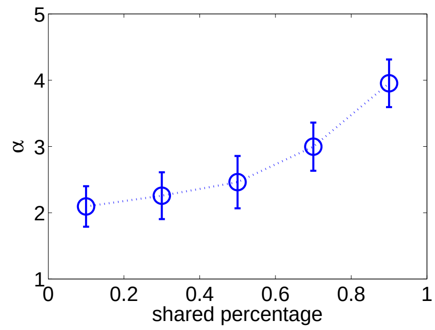
1.2 Missing value estimation

There usually exist missing values in real aCGH data, *e.g.* the data set from [1]. Our model can estimate missing values and we tested the estimation accuracy on the data set of chromosome 17 from [1]. We randomly picked part of observed entries for testing, labeled them as missing values and applied Algorithm 2. For comparison, we also estimated the testing entries by the nearest neighbor (NN) along each aCGH sample. The root-mean-squared error (RMSE) of the estimated values compared with their original values are given in Table 1. The RMSE of our method is consistently

*to whom correspondence should be addressed. Email: eexwzhou@ust.hk



(a)



(b)

Figure 1: (a) The distribution of selected parameters for a given data set, where the shared percentage is 0.5 and SNR = 1. (b) The selected α *vs.* the shared percentage. The result is averaged over 100 instances of synthesized data with SNR = 1.

Missing ratio	RMSE (TV-Sp)	RMSE (NN)
0.25	0.265	0.364
0.50	0.270	0.368
0.75	0.288	0.386
0.99	0.325	0.422

Table 1: Root-mean-squared error (RMSE) of missing value estimation. Two methods are tested: the TV-Spectral regularization used in this paper and the nearest neighbor (NN) method.

lower in Table 1 and it gets larger with the missing ratio increasing. When 99% of entries are missing, the RMSE of our method approaches the standard deviation of original data, which equals 0.324.

2 Proof of Theorem 1

In this document, we give a proof to Theorem 1 in the original paper. We follow the procedure similar to the convergence analysis of the *soft-impute* algorithm[2]. We use following definitions and notations.

Proximity operator Let r be a convex function. For each $\mathbf{x} \in \mathbb{R}^n$, the minimization problem

$$\min_{\mathbf{y}\in\mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + r(\mathbf{x}) \tag{1}$$

admits a unique solution denoted by $\mathrm{prox_r}(\mathbf{x})$. The mapping $\mathrm{prox_r}(\mathbf{x}) : \mathrm{\mathbb{R}^n} \to \mathrm{\mathbb{R}^n}$ thus defined is the proximity operator associated with function r [3].

Define the following function with variable $X \in \mathbb{R}^{m\times n}$

$$r_0(X) = \alpha \|X\|_* + \gamma \sum_{j=1}^{n} \|X_j\|_{TV}, \tag{2}$$

where α and γ are constants. Then, Equation (3) reads

$$\min_B \quad \frac{1}{2} \|D - B\|_F^2 + r_0(B). \tag{3}$$

Since $r_0(\cdot)$ is convex in $\mathbb{R}^{m\times n}$, the problem in Equation (3) admits a unique solution. We denote the solution by $\mathrm{prox}_{r_0}(D)$ by extending the definition of the proximity operator to $\mathbb{R}^{m\times n}$.

Define $f(B)$ to be the energy function of the extended model in Equation 4:

$$f(B) = \frac{1}{2} \|\mathcal{P}_\Omega(D) - \mathcal{P}_\Omega(B)\|_F^2 + r_0(B). \tag{4}$$

Then, Theorem 1 can be rephrased as:

Theorem 1 *The sequence* $\{B^k\}$ *generated by*

$$B^{k+1} = \mathrm{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^k)) \tag{5}$$

with any initial point B^0 *converges to a limit* B^∞ *that minimizes* $f(B)$.

Before we prove Theorem 1, we give several lemmas:

Lemma 1 *For any matrices* B *and* $\tilde{B}$, *define*

$$F(B|\tilde{B}) = \frac{1}{2} \|\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(\tilde{B}) - B\|_F^2 + r_0(B). \tag{6}$$

The sequence $\{B^k\}$ *defined in Equation (5) satisfies*

$$f(B^{k+1}) \leq F(B^{k+1}|B^k) \leq f(B^k). \tag{7}$$

Proof Note that $f(B) = F(B|B)$ for any B. Then we have

$$\begin{aligned}
f(B^k) =\ & F(B^k|B^k) \\
=\ & \frac{1}{2}\, \|\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^k) - B^k\|_F^2 + r_0(B^k) \\
\geq\ & \inf_B \frac{1}{2}\, \|\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^k) - B\|_F^2 + r_0(B) \\
=\ & F(B^{k+1}|B^k) \quad \text{(by definition } \; B^{k+1} = \arg\min_B \; F(B|B^k)) \\
=\ & \frac{1}{2}\, \|\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^k) - B^{k+1}\|_F^2 + r_0(B^{k+1}) \\
=\ & \frac{1}{2}\, \|\mathcal{P}_\Omega(D - B^{k+1}) + \mathcal{P}_{\Omega^\perp}(B^k - B^{k+1})\|_F^2 + r_0(B^{k+1}) \\
=\ & \frac{1}{2}\, \|\mathcal{P}_\Omega(D - B^{k+1})\|_F^2 + \frac{1}{2}\, \|\mathcal{P}_{\Omega^\perp}(B^k - B^{k+1})\|_F^2 + r_0(B^{k+1}) \\
\geq\ & \frac{1}{2}\, \|\mathcal{P}_\Omega(D - B^{k+1})\|_F^2 + r_0(B^{k+1}) \\
=\ & f(B^{k+1}) \quad \blacksquare
\end{aligned} \tag{8}$$

Lemma 2 *The proximity operator* $\mathrm{prox}_{r_0}(\cdot)$ *satisfies the following property for any* X_1 *and* X_2*:*

$$\left\|\mathrm{prox}_{r_0}(X_1) - \mathrm{prox}_{r_0}(X_2)\right\|_F^2 \leq \|X_1 - X_2\|_F^2\,. \tag{9}$$

Proof It has been proved that the proximity operator $\mathrm{prox}_\mathrm{r}(\cdot)$ of a convex and continuous function $r(\cdot)$ has the following nonexpansive property for any $\mathbf{x}_1$ and $\mathbf{x}_2$ [3, 4]:

$$\|\mathrm{prox}_\mathrm{r}(\mathbf{x}_1) - \mathrm{prox}_\mathrm{r}(\mathbf{x}_1)\|_2^2 \leq \|\mathbf{x}_1 - \mathbf{x}_1\|_2^2. \tag{10}$$

Since $r_0(\cdot)$ is convex and continuous, Equation (9) holds. $\blacksquare$

Lemma 3 *The sequence* $\{B^k\}$ *defined in Equation (5) satisfies:*

$$B^{k+1} - B^k \to 0 \quad as \quad k \to \infty. \tag{11}$$

Proof By definition we have

$$\begin{aligned}
\|B^{k+1} - B^k\|_F^2 &= \|\mathrm{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^k)) - \mathrm{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^{k-1}))\|_F^2 \\
\text{(by Lemma 2)} \;\; &\leq \|\left(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^k)\right) - \left(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^{k-1})\right)\|_F^2 \\
&= \|\mathcal{P}_{\Omega^\perp}(B^k - B^{k-1})\|_F^2 \\
&\leq \|B^k - B^{k-1}\|_F^2,
\end{aligned} \tag{12}$$

which means that

$$\|B^{k+1} - B^k\|_F^2 \leq \|\mathcal{P}_{\Omega^\perp}(B^k - B^{k-1})\|_F^2 \leq \|B^k - B^{k-1}\|_F^2. \tag{13}$$

Since the sequence $\|B^{k+1} - B^k\|_F^2$ is nonincreasing and lower-bounded,

$$\|B^{k+1} - B^k\|_F^2 - \|B^k - B^{k-1}\|_F^2 \to 0, \tag{14}$$

$$\|\mathcal{P}_{\Omega^\perp}(B^{k+1} - B^k)\|_F^2 - \|B^k - B^{k-1}\|_F^2 \to 0, \quad \text{as } k \to \infty. \tag{15}$$

Comparing Equation (14) with Equation (15), we have

$$\|\mathcal{P}_\Omega(B^{k+1} - B^k)\|_F^2 \to 0 \ \text{ as } k \to \infty. \tag{16}$$

Moreover, from Lemma 1 we have

$$F(B^{k+1}|B^k) - F(B^{k+1}|B^{k+1}) \to 0 \ \text{ as } k \to \infty, \tag{17}$$

which means that

$$\|\mathcal{P}_{\Omega^\perp}(B^{k+1} - B^k)\|_F^2 \to 0, \quad \text{as } k \to \infty. \tag{18}$$

Combining Equation (16) and Equation (18), we have

$$B^{k+1} - B^k \to 0 \ \text{ as } k \to \infty. \quad \blacksquare \tag{19}$$

Lemma 4 *The sequence $\{B^k\}$ defined in Equation (5) is bounded and every limit point B^* of the sequence satisfies:*

$$B^* = \text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*)), \tag{20}$$

$$f(B^*) = \min_B \ f(B). \tag{21}$$

Proof Since $f(B)$ is convex and the sequence $\{f(B^k)\}$ is bounded (by Lemma 1), the sequence $\{B^k\}$ is also bounded.

By the Bolzano-Weierstrass theorem, there exists a subsequence $\{n_k\} \in \{1, 2, \cdots\}$ such that

$$B^{n_k} \to B^* \ \text{ as } k \to \infty, \tag{22}$$

where B^* is a limit point.

By definition, we have

$$B^{n_k} = \text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^{n_k-1})). \tag{23}$$

Passing over to the limits on both sides of Equation (23), for the left side we have Equation (22), and for the right side we will show

$$\text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^{n_k-1})) \to \text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*)) \ \text{ as } k \to \infty. \tag{24}$$

To prove this, by Lemma 3 we have

$$\|B^{n_k} - B^{n_k-1}\|_F^2 \to 0 \ \text{ as } k \to \infty. \tag{25}$$

Since $B^{n_k} \to B^*$ as $k \to \infty$, we have

$$\|B^* - B^{n_k-1}\|_F^2 \to 0 \quad \text{as } k \to \infty. \tag{26}$$

From Lemma 2 and Equation (26), we have

$$\|\text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*)) - \text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^{n_k-1}))\|_F^2 \to 0, \tag{27}$$

which is equivalent to Equation (24).

Thus, passing over to the limits on both sides of Equation (23) we have

$$\begin{aligned} B^* &= \text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*)) \\ &= \arg\min_B \ \frac{1}{2} \ \|\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*) - B\|_F^2 + r_0(B) \end{aligned} \tag{28}$$

Due to the optimality condition of convex optimization [5],

$$0 \in -(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*) - B^*) + \partial r_0(B^*), \tag{29}$$

where ∂ denotes the subdifferential. That is,

$$0 \in -(\mathcal{P}_\Omega(D) - \mathcal{P}_\Omega(B^*)) + \partial r_0(B^*). \tag{30}$$

Note that

$$\partial f(B^*) = -(\mathcal{P}_\Omega(D) - \mathcal{P}_\Omega(B^*)) + \partial r_0(B^*). \tag{31}$$

Hence, B^* minimizes $f(B)$ which is convex. ∎

Proof of Theorem 1 Following Lemma 4, there exists a subsequence $\{n_k\} \in \{1, 2, \cdots\}$ such that $B^{n_k} \to B^*$ as $k \to \infty$, where B^* is a limit point. Since B^* satisfies Equation (20), we have

$$\begin{aligned} \|B^k - B^*\|_F^2 &= \|\text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^{k-1})) - \text{prox}_{r_0}(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*))\|_F^2 \\ &\leq \|(\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^{k-1})) - (\mathcal{P}_\Omega(D) + \mathcal{P}_{\Omega^\perp}(B^*))\|_F^2 \\ &= \|\mathcal{P}_{\Omega^\perp}(B^{k-1}) - \mathcal{P}_{\Omega^\perp}(B^*)\|_F^2 \\ &\leq \|B^{k-1} - B^*\|_F^2, \end{aligned} \tag{32}$$

which means that the sequence $\{\|B^k - B^*\|_F^2\}$ converges.

Since $B^{n_k} \to B^*$, for every positive real number ϵ there exists a natural number N_0 such that $\|B^{n_k} - B^*\|_F^2 < \epsilon$ when $n_k > N_0$. Suppose $n_k = N_1$ is the smallest n_k that makes $\|B^{n_k} - B^*\|_F^2 < \epsilon$. By Equation (32) we have $\|B^k - B^*\|_F^2 \leq \|B^{N_1} - B^*\|_F^2 < \epsilon$ when $k > N_1$. It means that, for every ϵ, we can also find a natural number N_1 such that $\|B^k - B^*\|_F^2 < \epsilon$ when $k > N_1$. Thus, the sequence $\{B^k\}$ converges to a limit point $B^\infty = B^*$. By Lemma 4, B^∞ minimizes $f(B)$. ∎

References

[1] J.R. Pollack, T. Sørlie, C.M. Perou, C.A. Rees, S.S. Jeffrey, P.E. Lonning, R. Tibshirani, D. Botstein, A.L. Børresen-Dale, and P.O. Brown. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proceedings of the National Academy of Sciences*, 99(20):12963, 2002.

[2] R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11:2287–2322, 2010.

[3] J.J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93(2):273–299, 1965.

[4] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877, 1976.

[5] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.